WORLD CLIMATE PROGRAMME

APPLICATIONS

# EXTREMES AND DESIGN VALUES IN CLIMATOLOGY

by

Tibor Faragó
Hungarian Meteorological Service, Budapest, Hungary

Richard W. Katz
National Center for Atmospheric Research*, Boulder, U.S.A.

(*the NCAR is sponsored by the National Science Foundation, U.S.A.)

WCAP - 14

WMO/TD-NO. 386

WORLD METEOROLOGICAL ORGANIZATION

(September, 1990)

The World Climate Programme launched by the World Meteorological Organization (WMO) includes four components:

- The World Climate Data Programme
- The World Climate Applications Programme
- The World Climate Impact Studies Programme
- The World Climate Research Programme

The World Climate Research Programme is jointly sponsored by the WMO and the International Council of Scientific Unions.

This report has been produced without editorial revision by the WMO Secretariat. It is not an official WMO publication and its distribution in this form does not imply endorsement by the Organization of the ideas expressed.

# EXTREMES AND DESIGN VALUES IN CLIMATOLOGY+


## SUMMARY

The analysis of extreme event statistics, including event probabilities, return periods and design values, for various climatological elements is one of the most important problems in practical applications in climatology. Typical climatological problems are related to extraordinarily high or low values of air temperature, high wind gusts, high or low precipitation amounts, high short-term rainfall intensities, and severe droughts or floods.

Quite a few statistical methods for extreme value analysis have been developed during the last 30-40 years. However, most of these techniques have not been widely utilized in applied climatology. The reasons for this neglect are many, including a lack of appreciation of the fundamental differences between statistical estimation for extremes and that for averages, the relative complexity of some of the techniques, and the lack of guidance concerning which of these methods is most appropriate for particular climatological problems.

The conceptual background of extreme value analysis is briefly reviewed, and then statistical methods for the estimation of extremal characteristics and design values for climatological elements are more extensively described. To provide practical guidance regarding the application of these statistical estimation techniques to extreme climatological events, several examples are also presented, and a computer program for the calculations involved is described.



+Tibor Faragó, Hungarian Meteorological Service, H-1525 Budapest, P.O. Box 38, Hungary

Richard W. Katz, National Center for Atmospheric Research*, CO-80307 Boulder, P.O. Box 3000, U.S.A. (*the NCAR is sponsored by the National Science Foundation, U.S.A.)

## 1. THE CONCEPT AND USE OF CLIMATOLOGICAL EXTREMES

Extreme climatic events are the objects of study for many purposes. The likelihood of their incidence needs to be properly taken into account in the design of various structures (e.g., buildings, power supply systems, dams) which are vulnerable to (or whose operation is dependent upon) the weather or climate. In particular, these structures must be able to withstand quite extreme conditions which might only rarely occur during their entire lifetime. The effectiveness of the economic production of various activities is also significantly influenced by extreme weather or climate events (e.g., crop production, water supply, navigation). Moreover, the extreme phenomena might act as a catalyst in alerting societies to their vulnerability to fluctuations or permanent changes in climate.

The main purpose of this report is to demonstrate the practical utility of some statistical methods for extreme value analysis. Typical meteorological elements for which design values or event probabilities are commonly desired include extremely high or low air temperatures, wind gusts, intense precipitation, and snow depths.

Extreme events can usually be defined in terms of unusual values of a sequence of observations of certain meteorological elements. The term "extreme events" is used in a broad sense, encompassing both the occurrence of extraordinary values (i.e., a record-breaking maximum or minimum) and the exceedance of (or falling below) a particular threshold level. Typically, the problem is to estimate the probability that an extreme value of a sequence of observations of a meteorological variable will be higher or lower than some constant threshold level, or alternatively, to estimate that threshold value which will be exceeded with a desired fixed, small probability. These extreme values and associated probabilities are then used in the solution of related design problems or cost-risk calculations.

Historical observations of the appropriate meteorological variables are used for the identification and fitting of the desired extremal probability distributions. The utility of these estimators depends to a great extent on the length and the homogeneity of the observational record, especially in cases when the return period of the required design value is significantly longer than the observational record. Homogeneity involves the assumption that no systematic changes of the underlying climatic conditions were occurring during the observational period; i.e., all extremal observations were taken from the same statistical population. In order to justify the extension of the results of extreme value analysis into the future, the same homogeneity assumption must also encompass the relevant future time period. More aspects of the use and the conceptual problems of the extreme values in climatology are treated, inter alia, by Hershfield (1962), WMO (1974), Page (1976), Hoyt (1981), IAEA (1981), Mearns et al. (1984), Antal et al. (1988).

## 2. THEORETICAL BACKGROUND OF EXTREME VALUE ANALYSIS

### THE DISTRIBUTION OF MAXIMA AND MINIMA

Since the first work began in the 1920's (Fisher and Tippett, 1928), the mathematical basis for extreme value analysis has rapidly developed (Gumbel, 1942, 1958; Jenkinson, 1955, 1969; Galambos, 1978; Leadbetter et al., 1983). In fact, some of the probability theory for extremes of stochastic processes was derived in almost complete generality over 40 years ago (e.g., Gnedenko, 1943). However, the treatment of issues such as statistical estimation and inference that are of practical importance in climate applications has lagged these theoretical developments.

Many questions regarding extreme climatic events concern either the estimation of the probability distribution or related characteristics of the maximum or minimum of a sequence of observations of a meteorological variable. Denoting by $V_i$ the value of the meteorological variable V for the i-th time period, let

$$X^{(m)}=\max\{V_1, V_2, \ldots, V_m\}, \quad X_{(m)}=\min\{V_1, V_2, \ldots, V_m\}.$$

In climatology, we are commonly interested in estimating the probability that the maximum value exceeds a certain threshold (i.e., $P\{X^{(m)}>x\}$), or analogously, the probability $P\{X_{(m)}<x\}$. If the observations of the meteorological variable are independent and identically distributed with the common distribution function $F(x)=P\{V_i\leq x\}$, then the exact distributions of the maximum and minimum can be simply expressed as

$$P\{X^{(m)}\leq x\} = [F(x)]^m \quad \text{and} \quad P\{X_{(m)}\leq x\} = 1-[1-F(x)]^m. \qquad (1)$$

The theory of extreme values (e.g., Leadbetter et al., 1983) has established that, for sufficiently large parent sample size m, the probability distribution of the standardized (or "reduced") maximum value $Y^{(m)}=(X^{(m)}-u_m)/b_m$, $b_m>0$, can be approximated by one of three possible forms of extremal distribution function:

| Gumbel asymptote | Fréchet asymptote | Weibull asymptote | |
|---|---|---|---|
| $G_1(y)=\exp(-e^{-y})$ | $G_2(y)=\exp(-y^{1/k})$ | $G_3(y)=\exp[-(-y)^{1/k}]$ | (2) |
| | $y>0,\ k<0$ | $y<0,\ k>0.$ | |

Similar formulae hold for the minima:

| | | | |
|---|---|---|---|
| $H_1(y)=1-\exp(-e^{y})$ | $H_2(y)=1-\exp[-(-y)^{1/k}]$ | $H_3(y)=1-\exp(-y^{1/k})$ | (3) |
| | $y<0,\ k<0$ | $y>0,\ k>0.$ | |

These different forms of asymptotic distribution arise depending on the shape of the tails of the probability distribution F(x) (its right-hand tail for the maximum, left-hand tail for the minimum). In practice, the sampling conditions (homogeneity, independence, sample size) affect the accuracy of approximations for extreme climate events based on this asymptotic theory.

The asymptotic extremal distributions involve three parameters; namely, k - the shape parameter, $u_m$ - the location parameter and $b_m$ - the scale parameter. Their determination depends upon the particular form of distribution F of the individual observations. The latter two parameters are also called the attraction coefficients.

The three different types of asymptotic distributions can be combined into one generalized form (Jenkinson, 1955, 1969):

$$G(y) = G(y;k) = \exp[-(1-k\cdot y)^{1/k}], \quad y<1/k. \tag{4}$$

In particular, this formula reduces to the Gumbel asymptote for k=0, the Fréchet asymptote for k<0, and the Weibull asymptote for k>0. The extremal types for the minimum can be expressed in an analogous generalized form.

Under suitable conditions, the estimation of the probability distribution of the maximum value, $p=P\{X^{(m)}\leq x\}$ can be derived from the corresponding values for the asymptotic extremal distribution:

$$P\{X^{(m)} \leq x\} \approx G(y;k), \quad y=(x-u_m)/b_m. \tag{5}$$

Obviously, an extreme value analysis for minimum values can easily be transformed into a problem for maximum values by multiplying the variables by -1, since $\max\{-V_i\}=-\min\{V_i\}=-X_{(m)}$. For this reason, further treatment will usually concern only the maxima, moreover, for sake of simplicity, the index m will be omitted wherever it is assumed to be given or fixed (i.e., $X=X^{(m)}$).

## QUANTILES, DESIGN VALUES, RETURN PERIODS

Instead of estimating the distribution of the maximum (or minimum), sometimes the inverse problem of determining a "design value" is required; that is, the value $x_p^{(m)}$ such that

$$P\{X^{(m)} \leq x_p^{(m)}\} = p. \tag{6}$$

In other words, $x_p^{(m)}$ is the p-th quantile of the extreme value distribution. Alternatively, it is convenient to convert the probability level of the design value $x_p$ into the "return period" $T = 1/(1-p)$. Here T represents the expected waiting time until the threshold $x_p$ is first exceeded, or the mean recurrence time between two such "threshold events". The concept is analogous for the minimum.

The design values can be easily expressed when the asymptotic extremal distributions are employed. The inverse formulae for the Gumbel and the generalized extreme distributions are given by

$$y_p=G_1^{-1}(p)=-\log(-\log p) \text{ and } y_p=G^{-1}(p)=[1-(-\log p)^k]/k, \tag{7}$$

respectively. Therefore, the estimated design value for the return period $T=(1-p)^{-1}$ years of the extremal variable X may be computed, given the attraction coefficients, from

$$x_p = b\, y_p + u, \tag{8}$$

where $y_p$ is also called the "reduced design value".

The principal question in the application of extreme-value theory is the accuracy of the approximation (5). In other words, it is the problem of the rate of convergence of the exact extremal distribution, $F^{(m)}$ say, to the asymptotic extremal distribution; or, in terms of more practical significance, the accuracy of the design value $x_p$ estimated from the asymptotic extremal distribution as compared to its actual (but usually unavailable) value $x_p^{(m)}$. The theoretical aspects of this convergence have been treated in a comprehensive way by Leadbetter et al. (1983). The problems of the finite (parent) samples have been considered, among others, by Tabony (1983) and Court (1986).

For relatively short parent series, the likelihood that the maximum $X^{(m)}$ will not represent the right-hand tail of the parent distribution is quite high. Some practical guidance is given by Court (1986) and Faragó et al. (1989) as regards the sample size m (and probability levels) for which the estimates based on the asymptotic extremal distribution are suitable. The use of the asymptotic extremal distribution results in systematically overestimated design values; for instance, $x_p^{(m)}=3.54$ for independent standard normal variates $V_i$ with m=50 and p=0.99, whereas the Gumbel asymptote produces $x_p=3.75$. Nevertheless, this error necessarily decreases to zero as m increases.

## FREQUENCY OF THRESHOLD EVENTS

In many applications, the frequency of occurrence of values exceeding (or falling below) a particular threshold is of interest. Formally, if m(x) denotes the number of observed values $V_i$ (i=1,2,...,m) that exceed some threshold x, then m(x) has an approximate Poisson distribution provided both m and x are sufficiently large. That is,

$$P\{m(x)=M\} \approx [\tau^M/M!] \exp\{-\tau\}, \qquad M=0,1,2,... \tag{9}$$

where $\tau = \tau(x) = m[1-F(x)]$ denotes the mean number of exceedances of the threshold x.

It is important to be aware that this Poisson approximation is mathematically equivalent to the limiting distributions for the maximum of a sequence mentioned earlier (as pointed out by Leadbetter et al., 1983). This equivalence arises through the dependence of the mean number of exceedances $\tau$ on the right-hand tail of the parent distribution F. In particular, when the probability of no exceedances is considered (i.e., M=0), the formula (9) reduces to the extremal distribution function (4), i.e.,

$$P\{m(x)=0\}=P\{V_1 \le x, V_2 \le x, \ldots, V_m \le x\}=P\{X^{(m)} \le x\}=\exp[-\tau(x)]. \qquad (10)$$

If the mean number of exceedances $\tau(x)=m[1-F(x)]$ has asymptotically the form

$$\tau(x)=\exp\{-(x-\mu)/\sigma)\}, \qquad (11)$$

then (10) becomes the Gumbel distribution function. The generalized extreme value distribution (4) arises analogously (Van Montfort and Witter, 1985). This "peaks over threshold" approach has been employed, for example, by Ross (1987) to extreme wind speeds.

## GENERALIZATIONS OF EXTREME VALUE THEORY

The classical theory of extreme values just presented has been developed for the case of a sequence of independent and identically distributed random variables. Of course, a sequence of observations of a particular meteorological variable does not ordinarily satisfy these assumptions. Nevertheless, the theory of extreme values can be extended to the case of dependent sequences (e.g., Leadbetter et al., 1983) to account for the persistence (or autocorrelation) possessed by most meteorological variables and to the case of non-identically distributed observations to account for the seasonality which is also inherent in most meteorological variables. In particular, the main results of extreme value theory can be applied to series of observations whose autocorrelations satisfy some meteorologically realistic conditions (e.g., "*m*-dependent" or Markovian stationary sequences for which the autocorrelations gradually decay to zero - a property which is typical for meteorological observations; Newell, 1964; Katz, 1977; Faragó, 1977; Romanenko, 1984; Katz, 1988). In some cases, scarcer sampling (including clustering of exceedances) is attempted to obtain more nearly independent samples (Graham, 1983; Buishand, 1985).

The extremal models for meteorological variables whose event occurrences are themselves a stochastic process (e.g., rainfall events or wind gusts) have been proposed by Revfeim (1983), Revfeim and Hessell (1984). Besides the common event size characteristics (e.g., rainfall amounts), the rate of event occurrences is also introduced as a parameter in such a random occurrence model.

The problems of the nonhomogeneity (or seasonality) in the observations and the extension of extreme value methodology to such observations have also been considered (Horowitz, 1980; Carter and Challenor, 1981; Revfeim, 1982; Challenor, 1983; Tabony, 1983). Despite these asymptotically valid theoretical results, significant deviations may be encountered in the calculation of the extremal characteristics for those - *finite*, *autocorrelated* and/or *nonhomogeneous* - samples which are typically used in climatology (Court, 1986; Faragó et al., 1989). Consequently, the methods described next are most readily usable for idealized (large, homogeneous, stationary) samples, and specialized treatment may be needed in the case of meteorological variables which have marked seasonality or strong dependence.

## 3. METHODS OF ESTIMATION OF EXTREMAL CHARACTERISTICS

### SAMPLE SERIES

*Parent and extreme samples*

The previous discussion has been explicitly concerned with the theoretical situation in which only one random sample from the parent distribution F is considered. In fact, several observed parent samples are commonly available, say n different samples with their corresponding maxima being denoted by

$$X_1 = X_1^{(m)} = \max\{V_{1,1}, V_{1,2}, \ldots, V_{1,m}\},$$
$$X_2 = X_2^{(m)} = \max\{V_{2,1}, V_{2,2}, \ldots, V_{2,m}\},$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$X_n = X_n^{(m)} = \max\{V_{n,1}, V_{n,2}, \ldots, V_{n,m}\}.$$

For the purposes of extreme value analysis, either these parent series are directly modelled (i.e., the parent distribution is estimated) with the subsequent choice of the corresponding extremal distribution, or more frequently, only the n maxima $X_1$, $X_2$, ..., $X_n$ are recorded and used for the estimation of the parameters $u_m$, $b_m$ and k of the proper extremal distribution. The former approach is the *only* option if just one record exists (i.e., n=1), for instance, in the case of the extremes for particular calendar days (say the maxima of daily mean temperatures on January 1). However, in most design problems, the daily observed values of a certain meteorological element as the parent variable and their annual maximum (or minimum) as the extreme variable are used (m=365 or 366). The size of the extreme sample of such annual values is typically n=30 to 100. Of course, other time periods are sometimes used, e.g., when monthly or seasonal extremes are to be evaluated. Most of the methods utilize the ordered sample, so henceforth, we assume (if otherwise not stated) that the sample elements of the maxima are already ranked in ascending order, $X_1 \le X_2 \le \ldots \le X_n$.

*Partial duration series and the series of large values*

A commonly used alternative sampling technique utilizes all of the observed values during the entire sampling period which exceed some high threshold value x:

$$V_{j,i}>x, \quad i=1,2,\ldots,m_j(x); \quad j=1,2,\ldots,n$$

(i.e., $m_j(x)$ observations exceed the level x during the j-th year, $0\leq m_j(x)\leq m$) so that the total number of peaks equals $\Sigma m_j(x)\leq n\cdot m$. These observations form the "partial duration series". As mentioned before, these threshold events follow an approximate Possion process under certain conditions for increasing threshold levels and sample sizes. (Analogously, for the analysis of the smallest values, those observations are selected which are below some threshold level.) In some cases, this sampling method leads to design value estimates which have smaller standard errors than those achieved from the series of the annual maxima. In other words, the partial duration series can preserve more information about the tail behaviour of the parent distribution which is especially essential when the extreme sample size n is relatively small. (Asymptotically, the two methods behave identically.) With modified annual and partial duration sampling techniques, either only a fixed number of the largest annual values (Sneyers and Vandienpenbeeck, 1983; Smith, 1986) or a given number of the exceedances (Buishand, 1989) is selected. Employing a fixed number of annual maxima or exceedances is a method which reduces the uncertainties of the design value estimates under certain conditions.

## ESTIMATION PROCEDURES

As already mentioned, the asymptotic distribution of the maximum or minimum is determined by the tail behavior of the parent distribution F. In meteorological applications, the most commonly used theoretical extremal distribution is the first or

Gumbel asymptote. This custom is in accord with the theoretical result that the maxima of variables whose parent distribution F follows the normal, lognormal, exponential, gamma or Weibull laws (which are extensively used, along with other distributions, for temperature, precipitation or wind speed variables) tends asymptotically to the Gumbel extremal distribution (e.g., Leadbetter et al., 1983).

In practice, however, both the number of observations m of the climatological variable over which the extreme value is being investigated and the number of these extreme sample values n are finite. When these sample sizes are too short, this fact often leads to uncertainties about the appropriate form of parent distribution F (especially in the tails where observations are necessarily rare) or about the direct selection of the extremal asymptote. In such cases, three different approaches might be considered: (i) the *exact* form of the extremal distribution for a finite number of observations is employed; (ii) the form of asymptotic extremal distribution is selected and fitted by empirical techniques; or (iii) distributions that do not arise in the theory of extremes are fit empirically.

We emphasize that these approaches can produce significant differences in the estimates of probabilities for rare events ordinarily of concern, or for the design values corresponding to relatively small probabilities. Because the derivation of the exact formulae is generally rather complicated (or even impossible to do explicitly), only approaches (ii) and (iii) are practical in climatological applications. Moreover, although alternative distributions, such as the lognormal distribution for maximum precipitation amounts, have been sometimes fitted, their lack of theoretical justification makes this approach questionable for large values of the parent sample size m (Buishand, 1986; Tiago de Oliveira, 1986; Sevruk and Geiger, 1981; Hershfield, 1962).

In the practical use of extreme value theory, two basic approaches may be followed. Either the hypothesis testing is accomplished first to select the proper extremal distribution type with a subsequent estimation of its parameters, or the generalized distribution (4) is fitted. The latter method, of course, will lead to better approximations, since it includes the former method as a special case.

## GRAPHICAL METHODS

As a first step in empirical procedures, graphical methods are used to get some insight into the behaviour of the extreme samples. These values are usually depicted on Gumbel probability paper whose vertical axis is the twice iterated logarithm; i.e., corresponding to the inverse of the Gumbel distribution function, so that extreme samples distributed according to the Gumbel law will lie on a straight line $y=(x-b)/u$, where $y$ is the reduced variate. Fréchet and Weibull samples will exhibit nonlinearity, curving away from this line for the large values downwards and upwards (Fig. 1). An analogous picture holds for minimum values; however, it is simpler to use these values multiplied by -1, thus reducing this problem again to the analysis of maximum values.

The coordinates of the points to be depicted on the Gumbel probability paper are $(X_j, p_j)$, $j=1,2,\ldots,n$, where the $X_j$'s are the sample elements (observed maximum values) ranked in increasing order and the $p_j$'s denote the empirical estimates of the values of their distribution function. The latter are generally presented by the median values, viz.,

$$p_j = (j-0.307)/(n+0.386) \approx (j-0.3)/(n+0.4). \qquad (12)$$

However, other estimators are sometimes employed (Jenkinson, 1969; Cunnane, 1978). These ordinates are called the *plotting positions*. Similar forms of extreme probability paper are available for the Fréchet and Weibull asymptotes (Sevruk and Geiger, 1981; IAEA, 1981).

## ESTIMATORS OF THE GENERALIZED EXTREME VALUE DISTRIBUTION

The direct application of the three-parameter model (4) is the most straightforward technique for extreme value analysis and the derivation of the related design values. Several simpler estimation procedures are available if the hypothesis of $k=0$, i.e., the validity of the Gumbel asymptote is accepted. This can be based on theoretical considerations (e.g., fitting such parent distributions which according to the extreme value theory converge to the Gumbel asymptote), or from the results of statistical hypothesis testing. In principle, these simplified algorithms are also applicable to the Fréchet or Weibull distribution because the transformations $\log(X_j-u)$ and $-\log(u-X_j)$ result in Gumbel variab-les when the $X_j$'s follow Fréchet and Weibull distribution laws, respectively. However, a first guess for the location parameter u is required to accomplish these transformations.

### *The maximum likelihood method*

The most complex and - in some respects - the most effective estimators are provided by the maximum likelihood method (Jenkinson, 1969). Its advantages and disadvantages versus the other three-parameter models (see below) have been thoroughly analyzed by Hosking et al. (1985). The estimates of the parameters of the extremal distribution are obtained by the solution of the maximum likelihood equations

$$[R-(P+Q)k^{-1}]k^{-1}=0, \quad (P+Q)k^{-1}b^{-1}=0, \quad Q\,b^{-1}=0, \qquad (13)$$

where $P=n-\Sigma e_j$, $Q=\Sigma e_j f_j-(1-k)\Sigma f_j$, $R=n-\Sigma y_j+\Sigma y_j e_j$ and $y_j=(x_j-u)/b$, $e_j=\exp(-y_j)$, and $f_j=\exp(ky_j)$. An iterative method (commonly, Newton-Raphson method) can provide reliable estimates for the parameters u, b and k; an effective computer program algorithm has been reported by Hosking et al. (1985). (The initial values for the iteration can be calculated, for instance, by the method of sextiles or by the method of two-parameter moments.)

*Method of sextiles*

With this statistical method (Jenkinson, 1969), the ordered sample of extremes is divided into six approximately equal parts. If $S_1$, $S_2$,..., $S_6$ are the empirical mean values of these sub-series, then the shape of the extremal distribution can be estimated by use of the statistic $D_s = (S_2-S_1)/(S_6-S_5)$ and the table (Jenkinson, 1969):

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $D_s$: | 0.08 | 0.11 | 0.16 | 0.23 | 0.32 | 0.43 | 0.58 | 0.79 | 1.05 | 1.39 | 1.82 | 2.38 |
| k : | -0.5 | -0.4 | -0.3 | -0.2 | -0.1 | 0.0 | +0.1 | +0.2 | +0.3 | +0.4 | +0.5 | +0.6 |
| $\mu_s$: | 1.54 | 1.22 | 0.99 | 0.82 | 0.69 | 0.58 | 0.49 | 0.41 | 0.34 | 0.28 | 0.23 | 0.18 |
| $\sigma_s$: | 2.85 | 2.24 | 1.83 | 1.55 | 1.34 | 1.20 | 1.09 | 1.01 | 0.96 | 0.92 | 0.89 | 0.88 |

The theoretical values of the sextiles mean $\mu$, and standard deviation $\sigma$, are used to estimate the attraction coefficients similarly to the empirical method of moments:

$$b = \sigma_s'/\sigma_s, \qquad u = \mu_s' - b\mu_s, \tag{14}$$

where $\mu_s'$ and $\sigma_s'$ are the empirical values derived from the six values of the sextiles sample $S_1, S_2, \ldots, S_6$.

*Probability weighted moments*

Employing the ordered sample series, the first three empirical probability weighted moments are obtained as $\mu_L = \Sigma X_j (p_j)^L$, $j=1,2,\ldots,n$; $L=0,1,2$ where $p_j=(j-0.35)/n$ are the "plotting positions" for the sample values. (Of course, the "exact" median formula (12) is also applicable, however, the former expression for the empirical probabilities produced the best results according to Hosking et al., 1985.) Then the estimates of the parameters of the generalized extreme value distribution are (Buishand, 1986):

$$k=7.8590+2.9554c^2, \text{ where } c=(2\mu_1-\mu_0)/(3\mu_2-\mu_0)-\log 2/\log 3$$
$$b=(2\mu_1-\mu_0)/[\Gamma(1+k)(1-2^{-k})], \quad u=\mu_0+b[\Gamma(1+k)-1]/k, \tag{15}$$

where $\Gamma(1+k)$ denotes the gamma function. This method has been widely investigated by Greenwood et al. (1979), Landwehr et al. (1979) and Hosking et al. (1985). In the latter paper, it was shown that for sample sizes $n<100$, this procedure gives the shape parameter estimation with the lowest standard deviation (compared to the methods of the maximum likelihood and the sextiles); however, its bias is generally larger than that for the other estimators.

## HYPOTHESIS TESTING

The choice of the proper extremal distribution type may be based on the methods of statistical hypothesis testing. This approach was common before the application of the three-parameter model (the generalized extreme value distribution) was popularized. It involves either a decision between two alternative types of asymptotes or, more frequently, between the Gumbel asymptote and the two other types. Actually, the latter is a test of hypothesis concerning the value of the shape parameter $k=0$ versus $k\neq0$. Such a decision is important because of the high sensitivity of the design values to the type selection.

To test $G_1$ against the two other types, the simplest statistic is as follows (Gumbel, 1965; Van Montfort and Gomes, 1985; Demarée and Sneyers, 1986):

$$D_{MED} = \log[(X_n-X^*)/(X^*-X_1)], \tag{16}$$

where $X_1\leq X_2\leq\ldots\leq X_n$ are the ordered extremal observations and $X^*$ denotes their median. It was proved (Gumbel, 1965) that $D_{MED}$ is asymptotically normally distributed (under the hypothetized condition that the $X_j$'s obey the Gumbel distribution) with mean and standard deviation

$$\mu(D_{MED}) = \log\{-\log(n)\cdot\log^{-1}[-\log 2\cdot\log^{-1}(1-0.5^{1/n})]\},$$
$$\sigma(D_{MED}) = [0.861 \log(n) - 0.490]^{-1}.$$

Therefore, at the 5% significance level, the null hypothesis of the Gumbel distribution is rejected if the standardized value $D_{MED}' = [D_{MED} - \mu(D_{MED})]/\sigma(D_{MED})$ lies outside the interval (-1.96,+1.96). It seems more reasonable to employ the test statistic $-D_{MED}$ whose sign coincides with the sign of the shape parameter k; then the decision is made in favour of the Fréchet distribution (with $k<0$) if $-D_{MED}'<-1.96$ or the Weibull type (with $k>0$) is accepted at the given significance level when $-D_{MED}'>+1.96$ .

Another test, that is locally most powerful, is determined by the log-likelihood function (Otten and Van Montfort, 1980; Tiago de Oliveira, 1986): $D_{ML3}=\Sigma L(Y_j)$, $L(y)=\delta[\log g(y;k)]/\delta k|_{k=0}= -y+(1-e^y)y^2/2$, where $g(y;k)$ is the density function of the generalized extreme-value distribution (4) for the reduced variate $y=(x-u)/b$. Replacing the parameters with their maximum likelihood estimators from (13), the empirical value of the above statistic is computed as:

$$D_{ML3} = 0.5 \, \{\Sigma(X_j-\mu)^2 - n\Sigma e_j(X_j-\mu)^2/\Sigma e_j\}\cdot b^{-2}, \qquad (17)$$

where $\mu=\mu(X)$ is the empirical mean of the extremal sample and $e_j=\exp(-X_j/b)$. The maximum likelihood test statistic $D_{ML3}$ is asymptotically normally distributed with zero mean and standard deviation $\sigma(D_{ML3})=(2.09797\cdot n)^{-1/2}$. One decides again in favour of the hypothesis $k=0$ or rejects it(at the significance level of 5%) depending on whether the standardized value $D_{ML3}'=D_{ML3}/\sigma(D_{ML3})$ is within or outside the interval (-1.96, +1.96). For the same reason as above, it is advisable to use the statistic $-D_{ML3}$.

The hypothesis test concerning the types of extremal asymptote may also be based on probability-weighted moments. Denoting by $D_{PWM}$ the corresponding estimator of the shape parame-ter k determined by (15), this variable also has an asymptotic normal distribution. It is asymptotically unbiased ($\mu(D_{PWM})$ tends to zero) and has standard deviation $\sigma(D_{PWM})=(0.5633/n)^{-1/2}$. Thus the standardized test statistic

$$D_{PWM}'=D_{PWM}/\sigma(D_{PWM}) \tag{18}$$

may also be employed in hypothesis testing for the type of extremal distribution (Hosking et al, 1985).

## PARAMETERS OF THE GUMBEL DISTRIBUTION

As mentioned before, several simplified estimators are available for the two parameters of the Gumbel distribution. This choice (i.e., the application of this asymptote) may be justified either on a theoretical basis or as a result of hypothesis testing. However, this problem is always solved for finite samples and it should be emphasized again that the application of the Gumbel distribution may lead to significant bias (i.e., overestimation) in the design values for large return periods.

### *Theoretical formulae*

When the asymptotic form of the extremal distribution has been established, its parameters may be determined from theoretical considerations. This is the only option when just one parent sample is available (e.g., as mentioned before, in the case of the observations for a particular calendar day). Because the attraction coefficients are only asymptotically unique, different choices are possible. For the first asymptote, Gumbel (1958) derived

$$\begin{aligned} &u_m=u^{(m)}:\ F(u^{(m)})=1-1/m, \quad b_m=[m\cdot f(u^{(m)})]^{-1} \text{ for the maxima,}\\ &u_m=u_{(m)}:\ F(u_{(m)})=1/m, \quad b_m=[m\cdot f(u_{(m)})]^{-1} \text{ for the minima,} \end{aligned} \tag{19}$$

where these location parameters are the corresponding quantiles of the distribution F(x) (and called also as the "characteristic extremes") and $f(x) = F'(x)$ is the parent density function. For the maxima, this location parameter is the value which on average reached or exceeded once out of m observations, since $m[1-F(u^{(m)})]=1$. The corresponding value of the density function (which is needed for the calculation of the scale parameter) may be derived or taken from a table if the parametric form of distribution function is known. For instance, if F is the exponential distribution (with unit scale parameter), then $b_m=1$ and $u_m=\log(m)$.

Other examples of sequences of the asymptotically valid parameters are deduced by Leadbetter et al. (1983): in particular, the attraction parameters for maximum values of standard normal variates are as follows:

$$b_m=[2\cdot\log(m)]^{-1/2}, \quad u_m=b_m-1-b_m[\log(\log(m))+\log(4\pi)]/2. \qquad (20)$$

For arbitrary (identically distributed and independent) normal variables $V_i$, $i=1,2,\dots,m$ with common mean value $\mu(V)$ and standard deviation $\sigma(V)$, the attraction coefficients of the maxima are determined from (20) as $b=\sigma(V)b_m$ and $u=\mu(V)+\sigma(V)u_m$. Note that in this case analytical expressions for the attraction coefficients could not be obtained from (19) directly, since F does not exist in closed form.

*Method of empirical moments*

Moment estimations are obtained using elementary properties of the Gumbel distribution function (e.g., Gumbel, 1958); namely, the expressions for its moments. The theoretical mean value and the standard deviation for this distribution (for the maximum variable $X^{(m)}$) are $\mu_m = u_m+\Gamma\cdot bm$, $\sigma_m = b_m\cdot \pi/\sqrt{6}$. Therefore, we have the following expressions

$$b_m = \sigma_m \sqrt{6}/\pi, \qquad u_m = \mu_m-\Gamma\cdot b_m, \qquad (21)$$

where the moments can be replaced by their empirical estimates being calculated from the sample $X_1, X_2, \ldots, X_n$ and $\Gamma=0.57722$ is the Euler's constant.

*Method of empirical reduced moments*

The mean value and the standard deviation for the reduced variable $Y=(X-u_m)/b_m$ (i.e., for the Gumbel distribution in (2)) are $\mu(Y)=\Gamma$ and $\sigma(Y)=\pi/\sqrt{6}$. These values can also be estimated empirically from the series $y_j=-\log(-\log(p_j))$ where $p_j$ are the plotting positions (12), $j=1,2,\ldots,n$. Then the parameters of the Gumbel's straight line (on the Gumbel probability paper) are empirically estimated as

$$b_m = \sigma_m/\sigma(Y), \qquad u_m = \mu_m - \mu(Y)b_m, \tag{22}$$

where $\mu_m=\mu(X^{(m)})$ and $\sigma_m=\sigma(X^{(m)})$ are empirically derived from the maximum sample. The empirical values of moments for the reduced variable depend only on the sample size n (not on the mean and standard deviation of the parent distribution F). This method was used, inter alia, by Sevruk and Geiger (1981).

*Two-parameter probability weighted moments*

In the singular case of zero shape parameter, the formulae (15) reduce to the following form (Greenwood et al.,1979):

$$b=(2\mu_1-\mu_0)/\log 2, \quad u=\mu_0-\Gamma\cdot b. \tag{23}$$

For the case of the Gumbel distribution, the efficiency of this method compared to the other "conventional" methods has been investigated by Landwehr et al. (1979). It was revealed that, unlike the method of moments or maximum likelihood, this procedure gives unbiased estimators of the attraction coefficients and the design values for independent observations. The efficiency (in terms of the variance of the estimators) of the quantile estimators from the probability-weighted moments is generally higher than that from the conventional moments.

*Estimates from the sample quantiles*

The simplest statistical estimators of the attraction coefficients are related to the particular quantiles of the given extremal asymptote. The Gumbel distribution can be easily evaluated at the values $y'=(x'-u)/b=0$ and $y''=(x''-u)/b=1$, namely, $p'=G_1(0)=e^{-1}=0.367$ and $p''=G_1(1)=\exp(-e^{-1})=0.692$, which provide just two simple equations for the unknown parameters u and b. Hence, the method of quantiles gives the estimates:

$$u=x_{0.367} \quad \text{and} \quad b=x_{0.692}-x_{0.367}, \tag{24}$$

where $x'=x_{0.367}$ and $x''=x_{0.692}$ are the $p'$- and $p''$-quantiles of the Gumbel distribution. These quantiles can be empirically found from the ordered extreme sample. Other quantiles would yield more efficient estimates to some extent (Tiago de Oliveira, 1986); however, the use of such estimators are acceptable only as first, simple and quick estimators of the attraction coefficients (which might be suitable for graphical purposes or for the initialization of more sophisticated iteration procedures).

*Linear estimators*

The linear order statistics for the parameters of the Gumbel extreme value distribution have been developed by Lieblein (1966, 1974). The coefficients of the best linear unbiased estimators,

$$u=\Sigma c_j X_j \quad \text{and} \quad b=\Sigma d_j X_j, \tag{25}$$

can be calculated depending on the sample size ($n\leq16$, $16\leq n<50$ and $50\leq n$) from the formulae and basic coefficient values tabulated in Lieblein (1974). Actually, the exact "best" coefficients $c_j$ and $d_j$ are used only for the small sample sizes ($n\leq16$); the use of a simplified procedure is recommended by Lieblein for larger sample sizes which also produces satisfactory results. The standard deviation of these linear estimators has only a relatively little improvement beyond n=10 over that for the optimal estimators (in the case of n=16, it is 99% of that of the theoretically best estimator for the location coefficient u, and 95% for the scale parameter b). In spite of their comparable effectiveness, these linear estimators are rarely used recently because other procedures (like the methods of moments or probability-weighted moments) with similar efficiency characteristics can be implemented much more easily.

*The two-parameter maximum likelihood method*

This method is a simplified version of the three-parameter maximum-likelihood procedure (13), with its implementation being relatively complicated in comparison with the above methods. An approximate version was deduced by Kimball in 1956 (Gumbel, 1958; Demarée and Sneyers, 1986), whereas the exact formulae were derived by Jenkinson (1969) and have been used, among others, by Tabony (1983), Boyack (1985) or Tiago de Oliveira (1986). The maximum likelihood equations are

$$u=-b\cdot\log[\Sigma\ e_j/n],\quad b=u-[\Sigma\ X_j e_j]/[\Sigma\ e_j],\quad j=1,2,\dots,n, \tag{26}$$

where $e_j=\exp(-X_j/b)$. Like the three-parameter method, an iterative method can be used for the solution where initialization is performed by either the method of moments (Tiago de Oliveria, 1986) or the sextile method (Jenkinson, 1969). The maximum likelihood estimators of u and b are asymptotically unbiased with approximate variances $\sigma^2(u) = b^2[1+6(1-\Gamma)^2\pi^{-2}]/n$ and $\sigma^2(b) = 6b^2\pi^{-2}/n$.

*Partial duration series and the Poisson model*

The parameter of the Poisson model $\tau(x)$ is just its theoretical mean or expected value. Its empirical estimate can be obtained from the ordinary partial duration sample in a simple manner; namely, as the mean number of "threshold events" or peaks $\tau'(x)=\Sigma m_j(x)/n$ $(j=1,2,..,n)$ where $m_j(x)$ denotes the number of exceedances of the threshold x from the j-th parent sample. The location parameter is estimated from (11), whilst the empirical scale parameter equals the mean deviation of the selected peaks from the value x, $V_{j,i}>x$ (Cunnane, 1973; Buishand, 1989)

$$u=x+b\cdot\log\tau'(x),\quad b=\Sigma_j\Sigma_i(V_{j,i}-x)/\tau'(x),\ j=1,2,...,n. \tag{27}$$

**Modifications to the method of partial duration statistics have been treated, inter alia, by Buishand (1989).**

## THE DESIGN VALUES AND THE ACCURACY OF THEIR ESTIMATES

Once the extreme asymptote is fitted and its parameters are estimated, the design values with the given return periods can readily be derived from (6)-(8). The most straigthforward way is to use the generalized formula (4) with the corresponding expression of the p-quantiles or T=1/(1-p) return period design values (8)

$$x_p=b\cdot[1-(-\log p)^k]/k+u,\ \text{if } k\neq0;\quad x_p=b\cdot[-\log(-\log p)]+u \text{ if } k=0,$$

where the parameters k, u and b are replaced with their empirical estimates. This is a point estimator, with its asymptotic accuracy being guaranteed by the relevant theoretical results (which hold under certain general conditions). However, their accuracy in practice is strongly dependent, inter alia, on the sample sizes and the probability level (return period or the threshold levels of the exceedance events). Consequently, it is also reasonable to consider confidence intervals for design values.

The accuracy of various statistical methods has been treated in several articles based on the asymptotic properties of the estimators and their empirical performance for simulated sample series (Greenwood et al., 1979; Landwehr et al., 1979; Hosking et al., 1985; Tiago de Oliveira, 1986). Usually, these estimators are asymptotically both unbiased and normally distributed, so that the approximate standard errors will readily determine the required interval estimates for either the parameters of the extremal distribution or for the design values. Specifically, the confidence intervals are expressed as $(x_p-\sigma_p,\ x_p+\sigma_p)$ at the approximate 68%-level, $(x_p-1.28\sigma_p,\ x_p+1.28\sigma_p)$ at the approximate 80%-level and $(x_p-1.65\sigma_p,\ x_p+1.65\sigma_p)$ at the approximate 90%-level, where $\sigma_p=\sigma\{x_p\}$ is the standard error of the estimator of the design value $x_p$. In particular, the three-parameter maximum likelihood method (13) and the cited computer program (Hosking, 1985) give also the estimated variances and covariances of the parameters; namely, the empirical estimates for $A=\sigma^2(u)$, $B=\sigma^2(b)$, $C=\sigma^2(k)$, $F=\mathrm{cov}(u,k)$, $G=\mathrm{cov}(b,k)$ and $H=\mathrm{cov}(u,b)$ from which the estimated variance of the p probability level design value becomes

$$\sigma^2\{x_p\}=A+B(y_p)^2+C(y_p')^2+2F\,y_p'+2G\,y_py_p'+2H\,y_p, \qquad (28)$$

where $y_p$ is the "reduced" design value for the generalized extremal distribution (7), $y_p=[1-(-\log p)^k]/k$, and $y_p'$ is its derivative, $y_p'=\delta y_p/\delta k=\{(-\log p)^k\cdot[1-\log((-\log p)^k)]-1\}/k^2$.

The formula is considerably simpler for the two-parameter maximum likelihood scheme (26) for the Gumbel asymptote:

$$\sigma^2\{x_p\}=\sigma^2(u)+\sigma^2(b)(y_p)^2+2\mathrm{cov}(u,b)y_p=[1+6(y_p+1-\Gamma)^2/\pi^2]b^2/n. \quad (29)$$

It is remarkable that the expression in parentheses depends only on the probability level or the return period $T=(1-p)^{-1}$. Thus, it may be calculated and tabulated independently of the particular sample; e.g., its value for T=100 is 4.05, so that $\sigma^2\{x_{0.99}\}=4.05\cdot b^2/n$. Obviously, the standard deviations of the quantile estimators can easily be obtained from those for the reduced variate, since $\sigma\{x_p\}=b\cdot\sigma\{y_p\}$. (Contrary to the above case, the partial duration series method (27) results in biased estimators of the design values.)

For small sample sizes, the accuracy of the estimators is much more sensitive to the sample properties. In the Gumbel case, according to the results produced for extensive simulated samples (Landwehr et al., 1979), the method of conventional moments (21) or the maximum likelihood procedure (26) result in considerably biased estimators, whilst the design values deduced from the probability-weighted moments are approximately unbiased for independent samples. For instance, their numerical experiments indicate that the $p=0.99$ probability-level reduced design value estimates $y_p'$ are less than the theoretical values $y_p=4.60$ for the sample size $n=29$ by 0.15 and 0.11 for the moment and the maximum likelihood methods, respectively. Of course, the smaller variance is obtained by the latter method; moreover, the variances of the estimators (and, in turn, the interval estimators) for the probability-weighted method are generally smaller than those for the ordinary moments. The standard deviations $\sigma(y_p)$ for the estimators from these two methods are about 10-20% larger than those from the maximum likelihood scheme for $p=0.99$ to 0.999, $n<1000$.

Similar numerical analysis has been performed for the generalized extreme-value distribution (4) by Hosking et al. (1985). Of the methods of sextiles (14), 3-parameter maximum-likelihood (13) and the probability-weighted moments (15), the latter produced the shape parameter estimators with generally the largest biases and the lowest standard deviations; however, this bias was almost negligible (not exceeding 10% of the value of the shape parameter) for moderate sample sizes ($n\geq50$). Finallyt, the estimators of the reduced design values from these methods, for instance, in the case of $k=+0.2$, $p=0.99$ ($y_p=3.01$) and $n=50$, have the biases of -0.01, -0.03 and +0.02, repectively. The biases and the standard deviations of estimators of $y_p$ from these methods are also comparable for such sample sizes. For small sample sizes ($n=15$ to 25), the use of the probability-weighted moments (or the method of sextiles) is usually more straightforward (i.e., much simpler and faster) compared to the maximum likelihood method and yields relatively small biases and, in general, substantially smaller standard deviations (and narrower interval estimators) of the design value estimators than those of the maximum likelihood estimators.

## 4. APPLICATIONS

Extreme value analysis, as well as the derivation of the design values, is a customary practice in applied climatological and other studies. These applications include the climatology for the construction of buildings (Page, 1976), problems related to intense rainfalls and the flood control (Jenkinson, 1969; Sevruk and Geiger, 1981), and accounting for extreme climatic events in the design of such special projects as, e.g., the nuclear power plants (IAEA, 1981).

The first type of problems encountered in the statistical application of extreme value analysis concerns the sampling conditions. A detailed discussion of this topic is beyond the scope of this review, and only a few aspects will be mentioned. The possibilities and limits of the generalization of extreme value theory to *dependent* (autocorrelated) and *nonstationary* series of parent observations was touched on above. The problem of the *finite* parent sample size, i.e., the effect of a limited number of parent variables m whose extreme value is considered on the adequacy of use of the extremal theory, is one of the most critical questions. Moreover, these attributes of the sample (finiteness, seasonality, etc.) are interrelated so that, for instance, in the case of the annual maximum temperature (taken from the daily maxima), roughly speaking, only the summer months contribute to the extremal distribution. There are several techniques in the literature regarding the treatment of these problems: applying mixed distributions, explicit parameterization of the seasonal cycle, removal of "outliers" and censored samples, clustering of the exceedances instead of considering all "threshold" events separately, more than one large value per year, etc. (Carter and Challenor, 1981; Challenor, 1982; Tabony, 1983; Sneyers and Vandienpenbeeck, 1983; Van Montfort and Gomes, 1985; Buishand, 1985, 1986; Faragó et al., 1989; Van Montfort, 1989).

Another source of occasionally significant uncertainties in the design value estimates is the selection of the statistical scheme and the interpretation of its results. In practical applications, generally only one particular method is employed: for example, point estimates might be derived by the method of moments, a priori accepting the Gumbel asymptote for the extremal variable. The illustrations below give some insight into the performance and comparability of the various methods for real series of observations. The following abbreviations are employed: MOM and MOM′ - method of moments with theoretical and empirical reduced characteristics (21) and (22); QNT - method of quantiles (24); LBL - linear estimates (25); PWM2 and PWM3 - method of 2- and 3-parameter probability weighted moments; ML2 and ML3 - 2- and 3-parameter maximum likelihood estimates.

*The "regular" case of the Gumbel asymptote*

As mentioned before, the Gumbel distribution is the extremal asymptote for the most typically used parent distributions in climatology. However, the extremes of *finite* number of observations deviate from this distribution law to a certain extent. Nevertheless, in some cases, the difference in the two- and the three-parameter approaches to the fitting of extremal distribution might be negligible. This situation is well illustrated by the sample of annual *maximum wind gusts* at Birmingham, Alabama, for the period 1944-1964, which was reported and analysed by Thom (1966). The original wind data should have been reduced to a "standard" height (of say 30 feet; this unit is retained for the comparability of the results) using the assumption of a logarithmic vertical wind profile. The statistical calculations with and without the hypothesis on the Gumbel distribution produce similar results (the unit of the location parameters and the design values are that of the wind speed in mph):

| | k | u | b | $x_{0.98}$ |
|---|---|---|---|---|
| PWM2 | 0 | 43.4 | 5.5 | 64.9 |
| PWM3 | *0.00* | 43.5 | 5.6 | 65.0 |
| ML2 | 0 | 43.9 | 4.5 | 61.3 |
| ML3 | *-0.05* | 43.8 | 4.4 | 62.6 |
| (SXL | *+0.04* | 43.7 | 4.7 | 60.8) |

The hypothesis testing perfectly supports the idea of accepting the Gumbel model (k=0), namely the standardized values of the test statistics (16-18) are as follows: -0.65, -0.15, and +0.02, respectively. Even in such a simplified case, depending on the requirements of the particular design problem, the differences in the estimates produced by the various two-parameter methods should be taken into account (Fig. 2). Therefore, it is more informative to use interval estimators. In the case of the maximum likelihood estimator (26), such an interval can be found employing the empirical standard deviation from (29), whose value is $\sigma(x_p)$=3.4 mph for p=0.98; consequently, the 80% and the 90% interval estimates of the 50-year return period design value read (56.9, 65.7) and (55.7, 67.0), respectively (in units of mph). (In his referenced work, Thom applied the Lieblein's linear estimator (25) directly to the logarithms of the reduced wind speed data under the assumption that the variable obeys the simplified Fréchet distribution with u=0. The calculations lead to a relatively small shape parameter, k=-1/11.74=-0.09, with $x_{0.98}$=61.3 mph which is rather close to the estimate $x_{0.98}$=62.03 mph derived for the Gumbel asymptote, k=0 from (25) using the linear scheme.)

A near zero empirical shape parameter has been found in many other cases. As another example, we illustrate the calculations for the annual *maximum daily precipitation amounts* at Bever (Switzerland), for the period 1901-70 (Sevruk and Geiger, 1981). The shape parameter estimates by the SXL, PWM3 and ML3 methods are 0.06, 0.07 and 0.07; the test statistics (16)-(18) give 0.68, 0.73 and 0.74. The various two-parameter methods produce the following parameter and design value estimates (mm):

| | k | u | b | $x_{0.98}$ | $x_{0.99}$ |
|---|---|---|---|---|---|
| MOM | 0 | 41.6 | 10.1 | 81.2 | 88.3 |
| LBL | 0 | 41.6 | 10.8 | 83.7 | 91.2 |
| PWM2 | 0 | 41.4 | 10.6 | 82.6 | 89.9 |
| ML2 | 0 | 41.5 | 10.7 | 83.3 | 90.8 |

For such a large sample size, there are only small differences in even the high return period design values. (In the referenced work, the modified method of moments (22) was used with parameter estimates u=41.4 mm and b=10.9 mm.) Some of these estimates are also presented in Fig.3.

In many cases, only a relatively short sample is available. Then the uncertainty about the point estimates is much higher and even more attention should be paid to the use of the interval estimates. Only a sample size of n=26 (annual data for 1940-65) was employed for the *maximum 24 h rainfall amounts* for the Chania-Kinakia catchment area (Kenya) by Jenkinson (1969). The Gumbel approach to these observations is again readily suitable (Fig.4), with shape parameter estimates from the SXL, PWM3 and ML3 methods 0.00, -0.06 and +0.05; and the test statistics (16)-(18) with empirical values of 0.09, 0.47 and -0.44, respectively. The estimates obtained for the Gumbel asymptote are (in inches):

| | k | u | b | $x_{0.98}$ | $x_{0.99}$ |
|---|---|---|---|---|---|
| MOM | 0 | 1.85 | 0.43 | 3.50 | 3.80 |
| LBL | 0 | 1.85 | 0.42 | 3.50 | 3.79 |
| PWM2 | 0 | 1.84 | 0.44 | 3.54 | 3.85 |
| ML2 | 0 | 1.85 | 0.43 | 3.53 | 3.83 |
| $\sigma(x_p)$ | | | | (0.30) | (0.34) |

The coefficient of variation for the 100-year design value from the ML2-method is 0.09, whilst this coefficient for the Birmingham wind data (for the same method and return period) is 0.06 . This coefficient is even higher for high return periods; e.g., $x_p$=5.82 inch, $\sigma(x_p)$=0.64 are the corresponding values for T=10000 years so that the coefficient of variation equals to 0.11 and, consequently, the 90%-level interval estimate is relatively wide, 5.82±1.06 inch. (Jenkinson (1969) used the ML2-technique for the same data and obtained the estimate $x_{0.9999}$=5.81 inch.)

*Sensitivity to model selection*

Large differences in estimated design values (based on the Gumbel and generalized extremal distributions) are obtained if the shape parameter k considerably deviates from zero. The principal question in the case of positive shape parameter values is whether to accept the smaller design values which are provided by the Weibull asymptote (k>0) or to use the Gumbel model anyway. The latter might cause severe overestimation. In this respect, there are two seemingly contradictory arguments: on the one hand, the Gumbel (and also the Fréchet) asymptote leads to infinitely increasing design values with increasing return periods which seems "physically" implausibe; on the other hand, in the case of high risks, it might be better to err on the 'safe' side.

The annual maximum temperatures at Budapest for the period 1921-1980 are considered. The results obtained by the different methods for the generalized and the Gumbel extremal distributions are listed below (°C):

| | k | u | b | $x_{0.98}$ | $x_{0.99}$ | $x_{0.995}$ | $x_{0.998}$ |
|---|---|---|---|---|---|---|---|
| SXL | *+0.24* | 34.4 | 1.9 | 39.3 | 39.8 | 40.2 | 40.7 |
| PWM3 | *+0.19* | 34.3 | 2.2 | 40.3 | 41.0 | 41.6 | 42.2 |
| ML3 | *+0.22* | 34.4 | 1.9 | 39.3 | 39.8 | 40.2 | 40.7 |
| MOM | 0 | 34.3 | 1.6 | 40.3 | 41.4 | 42.5 | 43.9 |
| MOM' | 0 | 34.2 | 1.7 | 40.8 | 42.0 | 43.1 | 44.7 |
| QNT | 0 | 34.4 | 1.4 | 39.9 | 40.8 | 41.8 | 43.1 |
| LBL | 0 | 34.4 | 1.7 | 41.0 | 42.2 | 43.4 | 44.9 |
| PWM2 | 0 | 34.1 | 1.9 | 41.4 | 42.7 | 44.0 | 45.7 |
| ML2 | 0 | 34.2 | 1.8 | 41.0 | 42.3 | 43.5 | 45.1 |

The various procedures yield results which are mainly deviating from each other between the 2- and 3-parameter methods. The sample maximum for the 60 years' period is 39.5°C which is in good agreement with the estimated design values for 50-100 years computed by the 3-parameter models. Acception of the Gumbel asymptote (k=0) would lead to considerably higher design values (Fig. 5). Although the standardized test statistics $-D_{MED}'=1.70$, $-D_{ML3}'=1.24$ and $D_{PWM}'=1.93$ do not give strong enough evidence to decide in favour of the Weibull asymptote (at the confidence level of 95%), it is essentially a philosophical issue of wether to employ (here and in general with the exception of practically small shape parameters) the Jenkinson formula (4) and the three-parameter methods. Although the 3-parameter methods are more complicated, they usually give significantly better fit for the observed samples and lessen the risk of deriving and applying serious overestimation for large return periods. In this particular case, a Weibull asymptote with the positive shape parameter $k \approx 0.22$ is selected and, in turn, the annual temperature maxima are asymptotically bounded from above by $u+b/k$, which gives 42.9°C as a point estimate with the ML3-estimates (being interpreted as the empirical mean value of the climatologically possible maximum temperature). Finally, it does not seem to matter which of the 3-parameter methods is applied: they lead to approximately similar results; however, the SXL- amd the PWM3-methods can be accomplished much more easily. More systematic analysis of the general performance of these methods on the basis of simulated sample series has been performed by Hosking (1985) and Hosking et al. (1985).

An analogous situation takes place for the annual *maximum flood* stage data for the Connecticut River at Hartford (1843-1934) investigated by Jenkinson (1969). However, the hypothesis on the first asymptote can be rejected on the basis of the conventional test statistics in this case. Otherwise, the Gumbel approach would again lead to serious overestimation (Fig.6). The test statistics produce empirical values of 2.85, 3.17 and 3.03, and the maximum likelihood estimates are as follows (in feet)

| | $k$ | $u$ | $b$ | $x_{0.98}$ | $x_{0.99}$ | $x_{0.995}$ | $x_{0.998}$ |
|---|---|---|---|---|---|---|---|
| ML2 | 0 | 19.2 | 3.4 | 32.3 | 34.7 | 37.0 | 40.1 |
| ML3 | +0.26 | 19.7 | 3.5 | 28.2 | 29.1 | 29.7 | 30.5 |

The sensitivity of the estimates to model selection is obvious and is also illustrated in Fig.6. (These values obtained by the ML3-method are the same as reported in the referenced work of Jenkinson (1969; p.205).)

### *The case of negative shape parameter*

Extreme variables with negative empirical shape parameters in climatological or hydrological practice are also customary. Notably, for such observations, the Gumbel type would yield apparent or real underestimation. Considering safety criteria, this approach (i.e., the case of $k<0$) has been recommended unconditionally for certain climatological elements (IAEA, 1981). According to Jenkinson (1969), the Fréchet model (k<0) leads to unrealistic, "fantastic" design values so that "common sense indicates that the negative k type should not be found in nature". Obviously, there is no upper limit for such estimates (with increasing return period). At least, the limited precision is clearly the case for such observations. The series of annual maxima of 60 minute rainfall intensities at Budapest, 1921-1988 is one such example (Fig. 7). The results of the maximum likelihood procedures are (mm/h):

| | k | u | b | $x_{0.98}$ | $x_{0.99}$ | $x_{0.995}$ | $x_{0.998}$ |
|---|---|---|---|---|---|---|---|
| ML2 | 0 | 16.8 | 8.2 | 48.8 | 54.6 | 60.3 | 67.8 |
| | | | σ(x) | 3.6 | 4.2 | 4.7 | 5.4 |
| ML3 | -0.15 | 16.2 | 7.7 | 56.9 | 67.0 | 78.1 | 94.7 |
| | | | σ(x) | 5.0 | 6.1 | 7.4 | 9.4 |

(The SXL- and the PWM3-method produce similar curves with shape parameters k=-0.16 .) The 60-year observed maximum is 57.0 mm/h. The 90% confidence intervals for the 500-year return period design value for the 2- and the 3-parameter models are (58.9,76.7) and (79.2,110.2).

### *Minimum values*

The easiest practical way to treat minimum data is to convert them to a series of maxima by multiplying each observation by -1, $V_i'=-V_i$. Then $\min\{V_i\}=-\max\{-V_i\}$ or $X_{(m)}=-X^{(m)\prime}$ Of course, the generalized formula for the minima $H(y)=1-\exp[-(1+ky)^{1/k}]$, $-1<ky$, is also applicable. To illustrate such an extreme value analysis, we employ the series of *winter temperature minima* at Saint Leo, Florida, 1932-1985. The conversion of the data does not affect the shape parameter. Applying this time only the relatively simpler methods, the following estimates are obtained (°F)

| | k | u | b | $x_{0.98}$ | $x_{0.99}$ | $x_{0.995}$ | $x_{0.998}$ |
|---|---|---|---|---|---|---|---|
| MOM | 0 | 29.6 | 3.7 | 15.3 | 12.7 | 10.2 | 6.8 |
| PWM2 | 0 | 29.6 | 3.6 | 15.5 | 13.0 | 10.5 | 7.2 |
| PWM3 | 0.22 | 29.1 | 4.2 | 18.0 | 16.9 | 15.9 | 14.8 |
| SXL | 0.23 | 29.3 | 4.5 | 17.7 | 16.5 | 15.5 | 14.4 |

The plots on the Gumbel probability paper for the minima (Fig.8) decrease towards the higher values; the Weibull type distribution estimate indicates that the minima are bounded from below. This time, the use of the Gumbel type would cause significant underestimation. The adequacy of the Poisson estimates (10) can also be shown: for this purpose, the average number of exceedances should be counted. For a threshold level of x=20°F, six events are found, which gives a Poisson parameter of $\tau(x)=0.11$ and $\exp\{-\tau(x)\}=0.90$; whilst, e.g., 0.94 is produced for this probability by the SXL-method. The problem with the direct application of the Possion scheme is that there are usually too few samples below (above in the case of maxima) the given threshold values; however, such low (high) levels are required to have approximately independent occurrences.

## REFERENCES


Antal E., Faragó T. and Glantz, M. H., 1988: On the concept of extreme meteorological and climatic events. Időjárás (J. of Hungarian Meteorol. Service) 92, 269-275.

Boyack, C. F., 1985: Investigation of the effect of length of record upon extreme values. Meteorol. Magazine, 114, 100-108.

Buishand, T. A., 1985: The effect of seasonal variation and serial correlation of the extreme value distribution of rainfall data. J. Clim. Appl. Meteor. 24, 154-160.

Buishand, T. A., 1986: Extreme value analysis of climatological data. Third Intern. Conf. on Statist. Climatology, Vienna, 145-158.

Buishand, T. A., 1989: The partial duration series method with a fixed number of peaks. J. Hydrol., 109, 1-9.

Carter, D. J. T. and Challenor, P. G., 1981: Estimating return values of environment parameters. Quart. J. Roy. Meteorol. Soc., 107, 252-266.

Challenor, P. G., 1982: A new distribution for annual extremes of environmental variables. Quart. J. Roy. Meteorol. Soc., 108, 975-980.

Court, A., 1986: Misapplications of extreme value statistics. Third Intern. Conf. on Statist. Climatology, Vienna, 218-220

Cunnane, C., 1973: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. J. Hydrol., 18, 257-271.

Cunnane, C., 1978: Unbiased plotting positions - a review. J. Hydrol. 37, 205-222.

Demarée, G. and Sneyers, R., 1986: The intensity-duration-frequency curves of the annual maximum rainfall at Uccle, Brussels. Third Intern. Conf. on Statist. Climatology, Vienna, 163-168.

Faragó T., 1977: On the estimation of probability distribution of maximum values and the statistical analysis of a wind velocity sample. Időjárás(J. of Hungarian Meteorol. Service) 81, 17-39.

Faragó T., Dobi I., Katz, R. W. and Matyasovszky I., 1989: Meteorological application of extreme value theory: problems of finite, dependent and non-homogeneous samples. Idöjárás (J. of Hungarian Meteorol. Service) 93, 261-275.
Fisher, R. A. and Tippett, L. H. C., 1928: Limiting forms of the largest or smallest members of a sample. Proc. Cambridge Philos. Soc., 24, 180-190.
Galambos, J., 1978: The asymptotic theory of extreme order statistics. Wiley, New York.
Gnedenko, B. V., 1943: Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. 44, 423-453.
Graham, A. E., 1983: A study of the Gumbel and Weibull methods of extreme value analysis using air temperature data from six Ocean Weather Stations. Meteorol. Magazine, 112, 303-317.
Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979: Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Res. Res. 15, 1049-1054.
Gumbel, E. J., 1942: On the frequency distribution of extreme values in meteorological data. Bull. Amer. Meteor. Soc. 23, 96-105.
Gumbel, E. J., 1958: Statistics of extremes. Columbia Univ. Press, New York.
Gumbel, E. J., 1965: A quick estimation of the parameters in Fréchet's distribution. Rev. Int. Stat. Inst., 33, 349-363.
Hershfield, D. M., 1962: An empirical comparison of the predictive value of three extreme-value procedures. J. Geophys. Res. 67, 1535-1542.
Hoyt, D. V., 1981: Weather "records" and climatic change. Clim. Change 3, 243-249.
Hosking, J. R. M., 1985: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. Royal Statist. Soc., 34, 301-310.
Hosking, J. R. M., Wallis, J. R. and Wood, E. F., 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27, 251-255.
IAEA, 1981: Extreme meteorological events in nuclear power plant siting, excluding tropical cyclones. IAEA Safety Guides, No. 50-SG-S11A, Vienna.
Jenkinson, A. F., 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quart. J. Roy. Meteor. Soc. 81, 158-171.
Jenkinson, A. F., 1969: Statistics of extremes. In "Estimation of Maximum Floods", WMO Techn. Note No.98, 183-257.
Katz, R. W., 1977: Precipitation as a chain-dependent process. J. Appl. Meteorol., 16, 671-676.
Katz, R. W., 1988: Statistics and decision-making for extreme meteorological events. Idöjárás (J. of Hungarian Meteorol. Service) 91, 23-35.
Landwehr, J.M., Matalas, N.C. and Wallis, J.R.,1979: Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Res. Res. 15, 1055-1064.
Leadbetter, M. R., Lindgren, G. and Rootzén, H., 1983: Extremes and related properties of random sequences and processes. Springer Verlag, New York.

Lieblein, J., 1966: A new method of analyzing extreme-value data. WMO T.N. No.81

Lieblein, J., 1974: Efficient methods of extreme-value methodology. NBSIR 74-602, National Bureau of Standards, Washington.

Mearns, L. O., Katz, R. W. and Schneider, S. H., 1984: Extreme high temperature events: changes in their probabilites with changes in mean temperature. J. Clim. Appl. Meteor. 23, 1601-1608.

Newell, G. F., 1964: Asymptotic extremes for m-dependent random variables. Ann. Math. Statist., 35, 1322-1325.

Otten, A. and Van Montfort, M. A. J., 1980: Maximum-likelihood estimation of the general extreme-value distribution parameters. J. Hydrol. 47, 187-192.

Page, J. K., 1976: Application of building climatology to problems of housing and building for human settlements. WMO Techn. Note No. 150

Revfeim, K. J. A., 1982: Seasonal patterns on extreme 1-hr rainfalls. Water Res. Res., 18, 1741-1744.

Revfeim, K. J. A., 1983: On the analysis of extreme rainfalls. J. Hydrol. 62, 107-117.

Revfeim, K. J. A. and Hessell, W. D., 1984: More realistic distributions for extreme wind gusts. Quart. J. Roy. Meteor. Soc., 110, 505-514.

Romanenko, T. P., 1984: Probability properties of maxima of time-series of wind speed modules (in Russian). Meteor. i Gidrol., No.2, 101-103.

Ross, W.H., 1987: A peaks-over-threshold analysis of extreme wind speeds. Canadian Journal of Statistics, 15, 328-335.

Sevruk, B. and Geiger, H., 1981: Selection of distribution type for extremes of precipitation. Oper. Hydrology Rep., WMO, No.15.

Smith, R.L., 1986: Extreme value theory based on the r largest annual events. J. Hydrol., 86, 27-43.

Sneyers, R. and Vandienpenbeeck, M., 1983: On the use of large values for the determination of the distribution of maximum values. Arch. Meteor. Geogh. Bioclim. B32, 279-286.

Tabony, R. C., 1983: Extreme value analysis in meteorology. Meteor. Magazine, 112, 77-98.

Thom, H. C. S., 1966: Some methods of climatological analysis. WMO Techn. Note No. 81

Tiago de Oliveira, J., 1986: Extreme values and meteorology. Theor. Appl. Clim. 37, 184-193.

Van Montfort, M. A. J. and Gomes, M. I., 1985: Statistical choice of extremal models for complete and censored data. J. Hydrol. 70, 77-87.

Van Montfort, M. A. J. and Witter, J.V., 1985: Testing exponentiality against generalized Pareto distribution. J. Hydrol., 78, 305-315.

Van Montfort, M. A. J., 1989: ML-estimation for EV1 and GEV (point, interval and censored data). Wageningen Agricultural Univ., TN-89-05

WMO, 1974: Guide to hydrological practices. WMO Publication No. 168

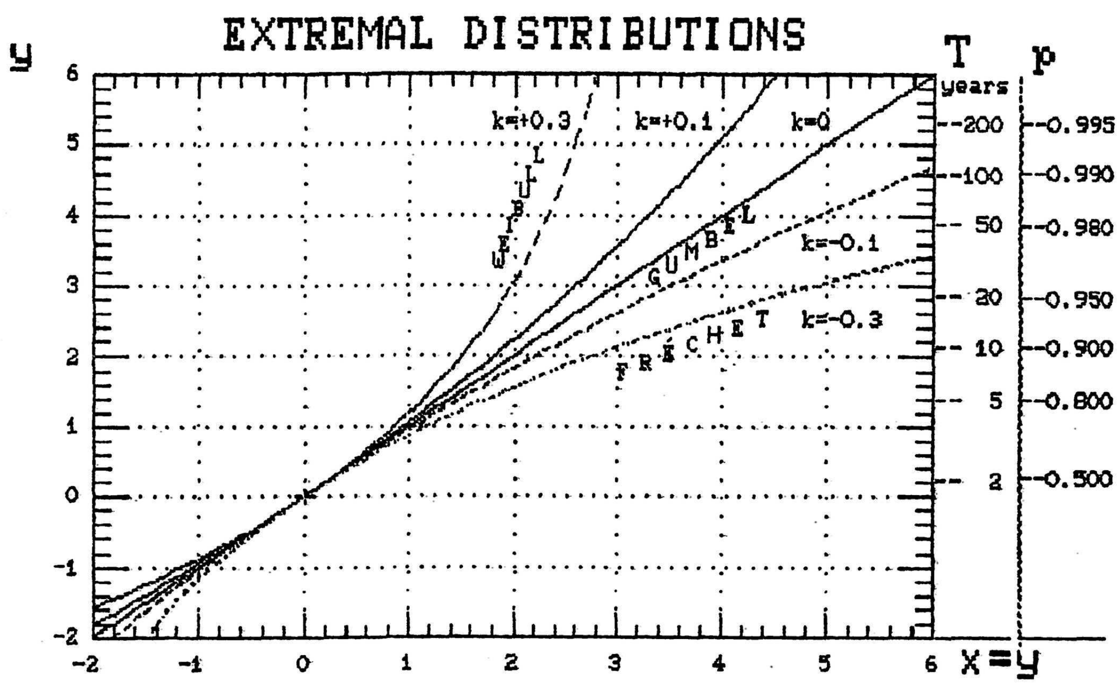


Figure 1

Types of asymptotic extremal distribution function depicted on Gumbel probability paper (Gumbel asymptote with shape parameter k=0, Weibull asymptote for k=0.1 and k=0.3, Fréchet asymptote with k=-0.1 and k=-0.3; y and T denote the reduced variable and the return period, respectively)

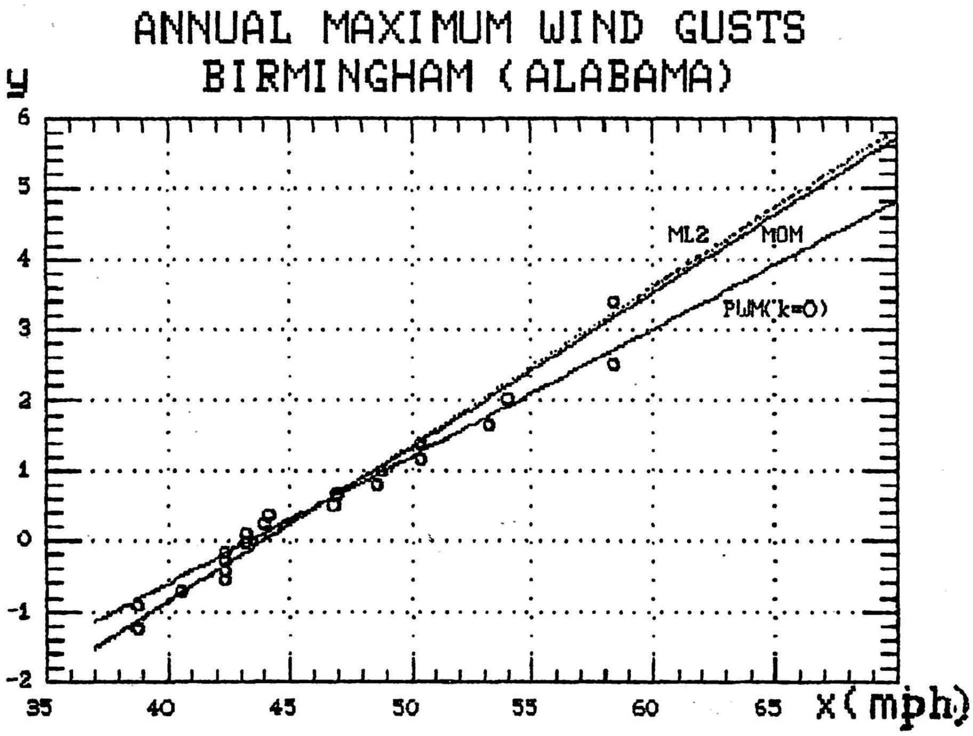


Figure 2

Estimators of the Gumbel extremal distribution for annual maximum wind gusts (mph) at Birmingham, Alabama, 1944-1964 (MOM - method of moments, PWM (k=0) - method of 2-parameter probability-weighted moments and ML2 - 2-parameter maximum likelihood estimator)

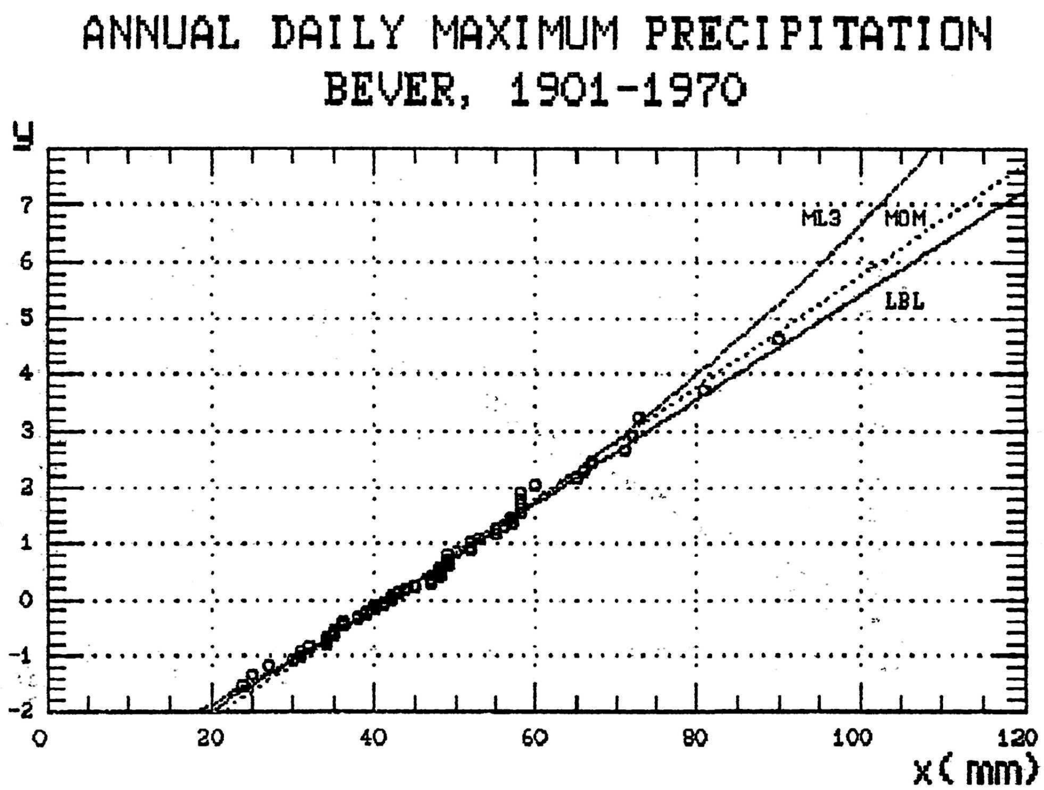


Figure 3

Extremal distribution functions fitted to the observations of the annual maxima of daily precipitation amounts (mm) at Bever, Switzerland, 1901-1970 (MOM - method of moments, LBL - Lieblein's linear estimator, ML3 - 3-parameter maximum likelihood estimator)

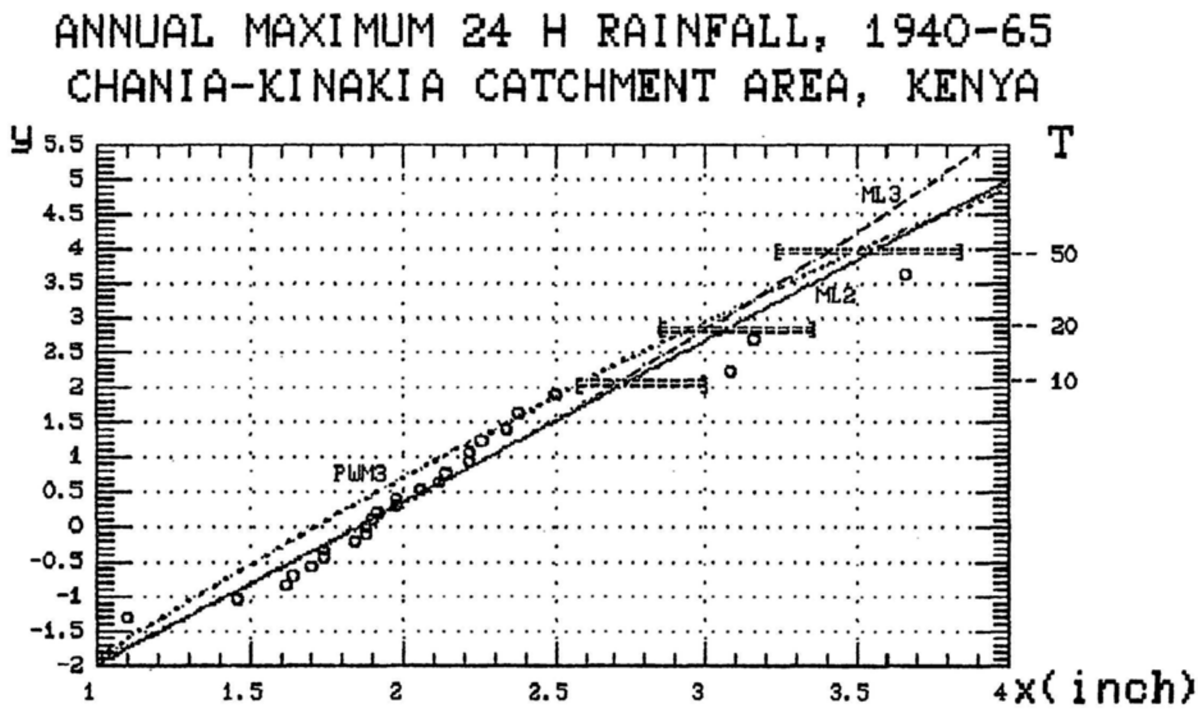


Figure 4

The results of extreme value analysis for series of annual maximum 24 h rainfall (inch) at Chania-Kinakia catchment area, Kenya, 1940-1965. Beside the curves derived by the 3-parameter probability-weighted moments (PWM3) and the 2- and 3-parameter maximum likelihood methods (ML2 and ML3), the [$-\sigma$ ,$+\sigma$ ] confidence intervals are also indicated for the ML2-estimators where $\sigma$ denotes the standard deviation of the estimated design values (with return periods 10, 20 and 50 years)

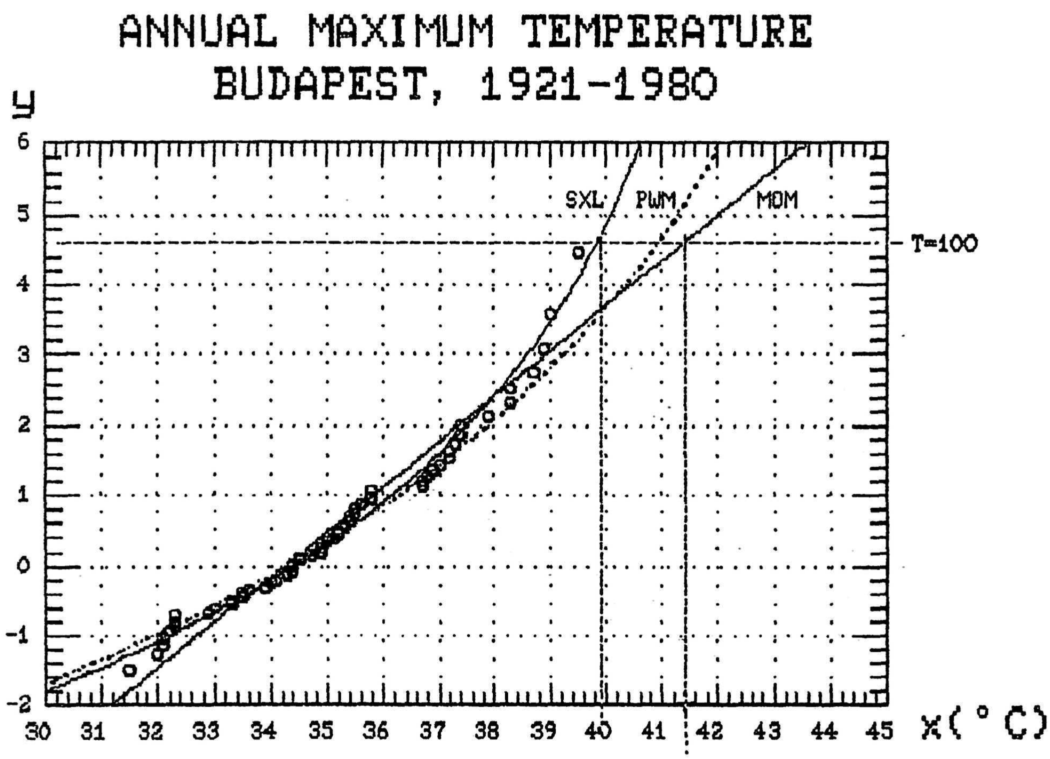


Figure 5

Plots of extremal distributions for annual maximum temperature observations at Budapest, 1921-1980, obtained by various methods (MOM - method of moments; SXL - method of sextiles; PWM - 3-parameter method of probability-weighted moments)

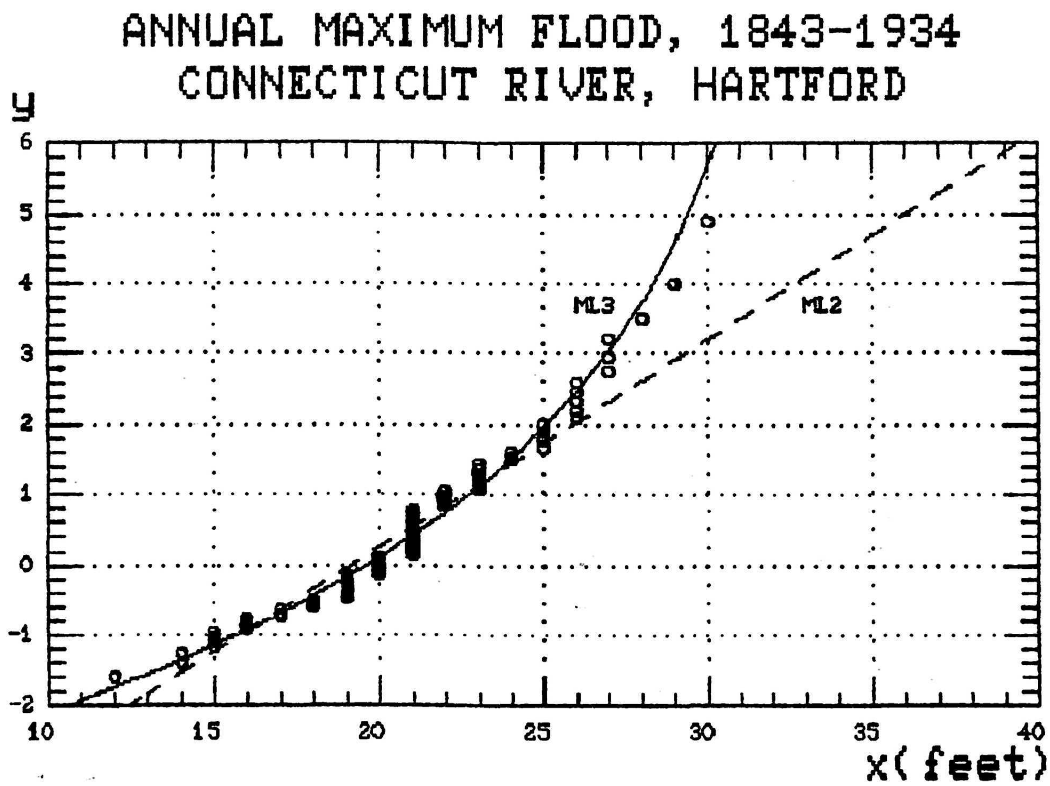


Figure 6

The observed values of maximum floods of the Connecticut River (feet) at Hartford, 1843-1934 and the approximate extremal distribution functions from the 2- and 3-parameter maximum likelihood estimators

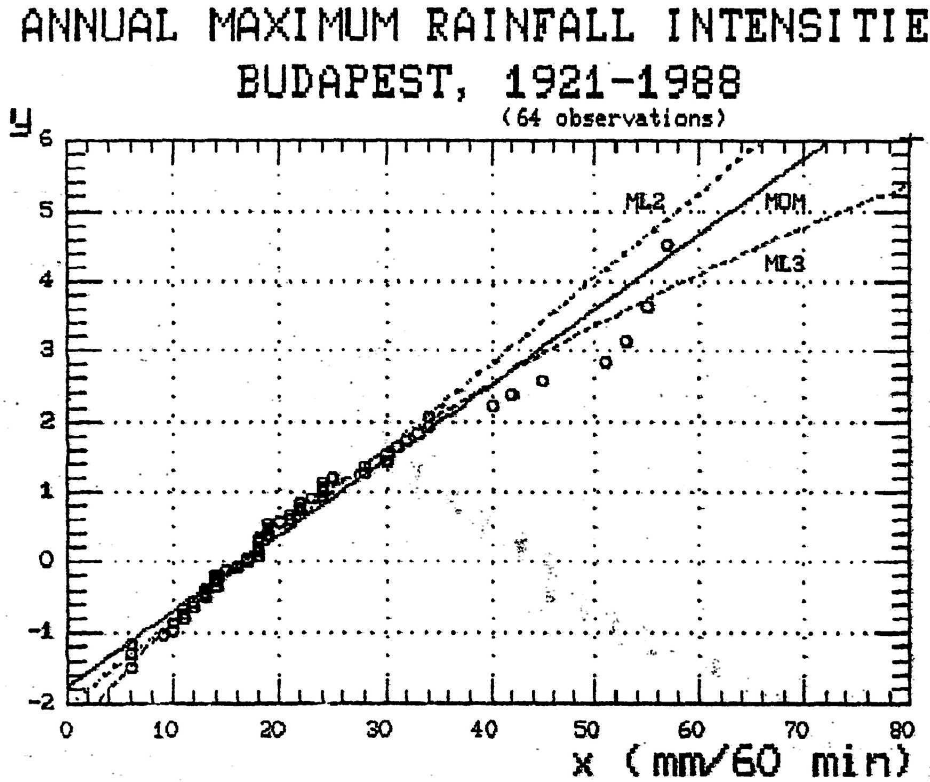


Figure 7

Plots of estimated extreme value distributions for the maxima of 60 minute precipitation intensities (mm/h) at Budapest, 1921-1988 (MOM - method of moments; ML2 and ML3 - the 2- and 3-parameter maximum likelihood methods)

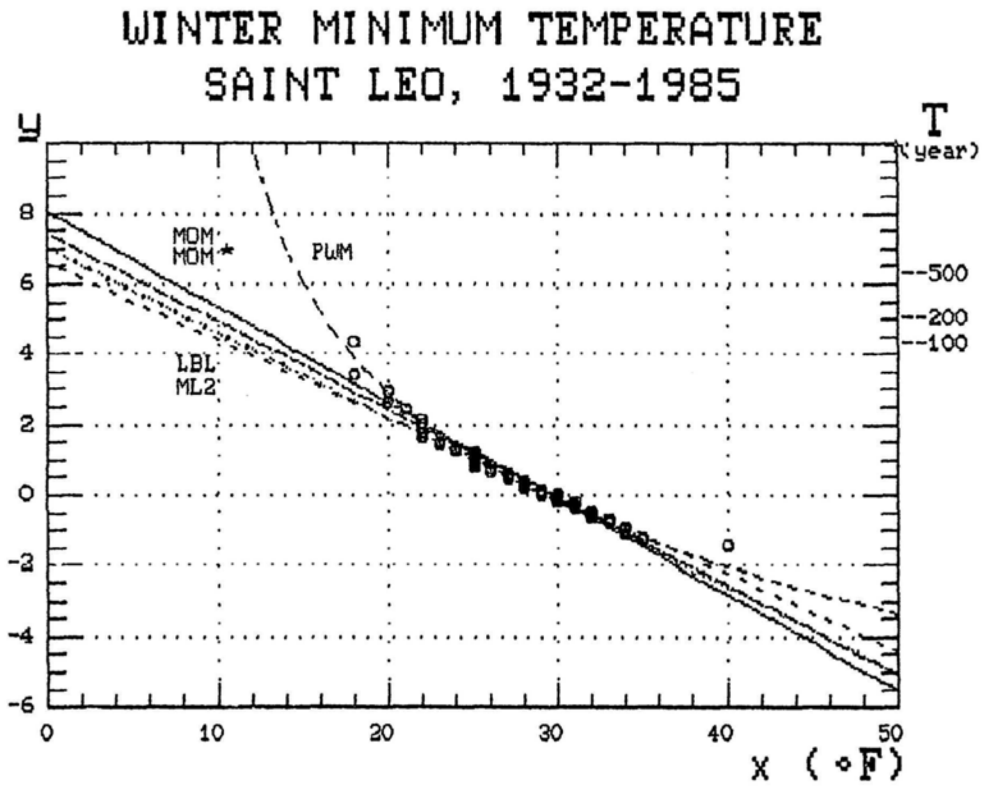


Figure 8

Annual minimum temperature observations ( °F) at Saint Leo, Florida, 1932-1985 and the plots of estimated extremal distribution function (four 2-parameter estimators: MOM and MOM* - ordinary and modified methods of moments, LBL - Lieblein's linear estimator and ML2 - the maximum likelihood method; PWM - method of the 3-parameter probability-weighted moments)

ANNEX

```
**************************************************
*               E X T R E M E S                  *
**************************************************
* Extremal  Characteristics,  Design  Values *
**************************************************
```

PROGRAMME "EXTREME.EXE" *

STATISTICAL METHODS OF EXTREME VALUE ANALYSIS

By: T. FARAGÓ and M. LAKATOS 1989/90
National Meteorological Service, Hungary
H-1525 Budapest P.O.Box 38

**PURPOSE**

COMPUTATION OF PARAMETERS OF EXTREMAL PROBABILITY DISTRIBUTION G(Y)=G(Y;K) (I.E., THE ATTRACTION COEFFICIENTS (U,B) AND THE SHAPE PARAMETER (K) OF THE GENERALIZED ASYMPTOTIC EXTREME-VALUE DISTRIBUTION FUNCTION) FROM A SAMPLE OF EXTREME VALUES (OBSERVATIONS), WHERE Y=(X-U)/B IS THE REDUCED VARIATE AND X IS THE OBSERVED (RANDOM) VARIATE

ESTIMATION OF THE DESIGN VALUES WITH VARIOUS RETURN PERIODS T (P-PROBABILITY LEVEL QUANTILES) OF THE EXTREMAL DISTRIBUTION

HYPOTHESIS TESTING FOR THE CHOICE AMONG THE EXTREME DISTRIBUTION TYPES TO ACCEPT OR REFUSE THE HYPOTHESIS ON THE GUMBEL DISTRIBUTION AGAINST THE WEIBULL OR THE FRECHET TYPES; ALTERNATIVELY, THE JENKINSON'S GENERALIZED DISTRIBUTION CAN BE USED: Prob{Y<C}=G(Y)=EXP(-(1-K*Y)^(1/K)), K*Y<1

---

* Run on XT/AT with Hercules monitor

METHODS FOR PARAMETER ESTIMATES (REALIZED IN SUBROUTINES):

2-PARAMETER CASE (FOR GUMBEL DISTRIBUTION WITH K=0)

MOM : METHOD OF MOMENTS
ANA : METHOD OF "ANALITICAL" (THEORETICAL) COEFFICIENTS
QNT : METHOD OF EMPIRICAL QUANTILES
LBL : LIEBLEIN'S METHOD ("BLUE")
PWM : METHOD OF PROBABILITY-WEIGHTED MOMENTS
ML2 : MAXIMUM-LIKELIHOOD ESTIMATION

3-PARAMETER CASE (JENKINSON'S GENERALIZED DISTRIBUTION)

SXL : METHOD OF SEXTILES
PWM : METHOD OF PROBABILITY-WEIGHTED MOMENTS
ML3 : MAXIMUM-LIKELIHOOD ESTIMATION

TEST : HYPOTHESIS TESTS

INPUT DATA

FILE OF SERIES OF EXTREMAL SAMPLES (E.G., SERIES OF ANNUAL TEMPERATURE MAXIMA OR MINIMA); ASCII-FILE OF DATA SEPARATED BY DELIMITERS (E.G., EACH VALUE IN A SEPARATE LINE); (ALTERNATIVELY, A RANDOM SEQUENCE OR AN ILLUSTRATIVE SAMPLE SERIES CAN BE USED).

OUTPUT

TABLES OF ESTIMATES OF PARAMETERS OF EXTREMAL DISTRIBUTION AND THE DESIGN VALUES; LISTS OF ORDERED EXTREME SAMPLE VALUES, REDUCED VALUES, EMPIRICAL PROBABILITIES (OPTIONAL).
THE OUTPUT RESULTS CAN BE SENT TO THE SCREEN, SAVED IN A FILE OR PRINTED.

PLOTTING

THE OUTPUT RESULTS CAN ALSO BE USED TO PLOT THE EMPIRICAL AND THE FITTED THEORETICAL DISTRIBUTIONS ON A GUMBEL PROBABILITY PAPER. THE EMPIRICAL PROBABILITIES (PLOTING POSITIONS) AND THE VALUES OF THE REDUCED VARIATE ARE LISTED IN THE TABLE OF EMPIRICAL PROBABILITIES. SCHEMATICALLY:

```
  Y 6-+----+----+----+----+----+----+----+----+ T     PROB
      ¦ EXDEMO.DAT                            ¦200    0.995
    5-¦- - - - - - - - - - - - - - - - - - - -¦
      ¦                            o          ¦100    0.990
    4-¦- - - - - - - - - - - - - - - - - - - -¦ 50    0.980
      ¦                           o           ¦
    3-¦- - - - - - - - - - - - - oo- - - - - -¦ 20    0.950
      ¦                      o                ¦
    2-¦- - - - - - - - * - * - - - - - - - - -¦
      ¦               *                       ¦
    1-¦- - - - - - *** - - - - - - - - - - - -¦
      ¦         ***                           ¦   * ~more sample
    0 ¦- - - **- - - - - - - - - - - - - - - -¦           points
      ¦     **                                ¦   o ~one sample
   -1 ¦- * * - - - - - - - - - - - - - - - - -¦           point)
      +--*-+----+----+----+----+----+----+----+
      0        20        40        60        80 mm/60 min
```

PROGRAMME SEGMENTS (IN SOURCE FORM *.FOR; AS OBJECT MODULES *.OBJ)

EXMAIN: MAIN PROGRAMME
EXDATA: DATA INPUT
EXSIMU: SIMULATION OF DATA
EXSTAT: BASIC STATISTICS
EXHEAD: HEAD OF OUTPUT TABLES
EXDESG: DESIGN VALUES
EXLIST: EMPIRICAL PROBABILITES

EXX2ANA: ANALYTICAL COEFFICIENTS
EXX2LBL: LIEBLEIN'S METHOD
EXX2ML2: 2-PARAM. MAXLIKE METHOD
EXX2MOM: METHOD OF MOMENTS
EXX2QNT: METHOD OF QUANTILES
EXX3SXL: METHOD OF SEXTILES
EXX3PWM: PROB.WEIGHTED MOMENTS
EXX3ML3: 3-PARAM. MAXLIKE METHOD
EXXTEST: HYPOTHESIS TESTING

ADDITIONAL FILES

README.BAT   EXDEMO.DAT   EXLINK.BAT
EXTREME.DOC   EXLINK.SGM

EXX2LBL.DAT: CONSTANTS OF EXX2LBL
EXSAFE.OBJ : I/O-ROUTINES

THE SOURCE PROGRAMMES ARE WRITTEN IN FORTRAN (F77), TRANSLATED BY FORTRAN* TRANSLATOR AND LINKED* UNDER DOS (VERSIONS 3.30). THE PROGRAMME SEGMENTS ARE GIVEN IN SOURCE FORM (*.FOR) AND AS OBJECT MODULES (*.OBJ).
IN THE CASE OF THE MODIFICATION OF ONE OR MORE SEGMENTS, AFTER THEIR RETRANSLATION, ALL OBJECT MODULES SHOULD BE LINKED AGAIN. FOR THIS PURPOSE, THE EXLINK.BAT BATCH FILE (WITH PARAMETERS IN EXLINK.SGM) CAN BE USED.

---

* See "README.BAT"

REFERENCES: Jenkinson, A.F., 1969: Statistics of extremes. In "Estimation of maximum floods", WMO, T.N. No.98; Gumbel, E.J., 1958: Statistics of extremes. Columbia Univ. Press, N.Y.; Hosking, J.R.M., 1985: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. Royal Statist. Soc., 34, 301-310.; Leadbetter, M.R., G. Lindgren and H. Rootzen, 1983: Extremes and related properties of random sequences and processes. Springer-Verlag, N.Y.; Lieblein, J., 1974: Efficient methods of extreme-value methodology. NBSIR 74-602, Nat'l Bureau of Standards, Washington; Otten, A. and Van Montfort, M.A.J., 1980: Maximum-likelihood estimation of the general extreme-value distribution parameters. J. Hydrol., 47, 187-192.; Tabony, R.C., 1983: Extreme value analysis in meteorology. Meteor. Mag., 112, 77-98.; Tiago de Oliveira,J., 1986: Extreme values and meteorology. Theor. and Appl. Climat.,37, 184-193.; Farago T., Dobi I., R.W. Katz and Matyasovszky I., 1989:Meteorological application of extreme value theory. Idojaras (J.of Hungarian Meteor. Service), 93, 261-275.; Int'l Atomic Energy Agency, 1981: Extreme meteorological events in Nuclear Power Plant Siting.Vienna(No.50-SG-S11A) Buishand, T.A., 1985: The effect of seasonal variation and serial correlation on the extreme value distribution of rainfall data.J.Climate Appl. Meteor., 24, 154-160.; Sevruk, B. and H. Geiger, 1981: Selection of distribution types for extremes of precipitation.Oper.Hydrol.Rep.,WMO,No.15.

NOTE: Copies of the software (on diskettes useable for IBM-compatible PCs) may be obtained from the authors at the address given on the first page of this annex.

Example of output

```
****************EXTREME VALUE ANALYSIS***********
*************************************************
ANNUAL MAX 24 H RAINFALL AT C-K CATCHMENT, KENYA, 1940-65 (INCH)
FILE NAME: MAXPKENY.DAT     <MAXIMA>       EXTREME SAMPLE SIZE=   26
MEAN=    2.09  STAND.DEVIATION=      .55   MAX=    3.65  MIN=    1.09
****************2-PARAMETER METHODS (  GUMBEL )
*************************************************
                return period        10        20        50       100       200       500      1000      2000      5000     10000
METHOD OF MOMENTS (THEORETICAL)
parameters of the extremal distribution:
K=   .00     U=    1.85    B=     .43
                design values:     2.80      3.11      3.50      3.80      4.10      4.49      4.78      5.08      5.47      5.76
METHOD OF MOMENTS (EMPIRICAL)
parameters of the extremal distribution:
K=   .00     U=    1.84    B=     .46
                design values:     2.97      3.20      3.63      3.95      4.26      4.68      5.00      5.32      5.74      6.06
METHOD OF QUANTILES
parameters of the extremal distribution:
K=   .00     U=    1.88    B=     .33
                design values:     2.62      2.86      3.17      3.40      3.63      3.93      4.16      4.39      4.69      4.92
LINEAR UNBIASED ESTIMATES (LIEBLEIN)
parameters of the extremal distribution:
K=   .00     U=    1.85    B=     .42
                design values:     2.80      3.10      3.50      3.79      4.08      4.47      4.76      5.06      5.44      5.74
METHOD OF PROBABILITY-WEIGHTED MOMENTS
parameters of the extremal distribution:
K=   .00     U=    1.84    B=     .44
                design values:     2.82      3.14      3.54      3.85      4.15      4.55      4.86      5.16      5.56      5.86
MAXIMUM LIKELIHOOD METHOD
parameters of the extremal distribution:
K=   .00     U=    1.85    B=     .43
          sd(U)=     .09 sd(B)=    .07
                design values:     2.82      3.13      3.53      3.83      4.13      4.53      4.83      5.13      5.52      5.82
          standard deviations:      .20       .24       .30       .34       .39       .45       .49       .54       .59       .64
****************3-PARAMETER METHODS (JENKINSON)
*************************************************
                return period        10        20        50       100       200       500      1000      2000      5000     10000
METHOD OF SEXTILES
sextile ratio=    .434
parameters of the extremal distribution:
K=   .00     U=    1.84    B=     .39
                design values:     2.72      3.00      3.37      3.65      3.92      4.28      4.55      4.83      5.19      5.46
METHOD OF PROBABILITY-WEIGHTED MOMENTS
parameters of the extremal distribution:
K=  -.06     U=    1.71    B=     .40
                design values:     2.68      3.02      3.49      3.86      4.25      4.78      5.21      5.66      6.28      6.77
MAXIMUM LIKELIHOOD METHOD
parameters of the extremal distribution:
K=   .05     U=    1.86    B=     .43
sd(K)= .11 sd(U)=     .09 sd(B)=    .06
                design values:     2.79      3.07      3.41      3.66      3.90      4.20      4.43      4.64      4.92      5.12
          standard deviations:      .27       .44       .72       .98      1.27      1.71      2.07      2.46      3.01      3.46
****************HYPOTHESIS TESTING**************
*************************************************
    MAX-LIKE TEST              MEDIAN TEST            PROB-WEIGHTED TEST
      -VN=    .465              -GN=    .087              ZN=   -.439
```

## REPORTS PUBLISHED IN THE WORLD CLIMATE APPLICATIONS PROGRAMME SERIES

| | |
|---|---|
| WCAP - 1 | CLIMATE AND HUMAN HEALTH. Proceedings of the Symposium in Leningrad, 22-26 September 1986, Volume I |
| WCAP - 2 | CLIMATE AND HUMAN HEALTH. Proceedings of the Symposium in Leningrad, 22-26 September 1986, Volume II |
| WCAP - 3 | ANALYZING LONG TIME SERIES OF HYDROLOGICAL DATA WITH RESPECT TO CLIMATE VARIABILITY - Project Description |
| WCAP - 4 | WATER RESOURCES AND CLIMATIC CHANGE: SENSITIVITY OF WATER-RESOURCE SYSTEMS TO CLIMATE CHANGE AND VARIABILITY. Norwich, U.K., November 1987 |
| WCAP - 5 | FOURTH PLANNING MEETING ON WORLD CLIMATE PROGRAMME - WATER. Paris, 12-16 September 1988 |
| WCAP - 6 | CLIMATE APPLICATIONS: ON USER REQUIREMENTS AND NEED FOR DEVELOPMENT [Reports of the CCl rapporteurs on Users' Requirements and Publicity (F. Singleton) and New Approaches in Applications (D.W. Philips) to the tenth session of the Commission for Climatology, Lisbon, April 1989] |
| WCAP - 7 | DROUGHT AND DESERTIFICATION. [Report of the CCl Rapporteur on Drought and Desertification in Warm Climates to the tenth session of the Commission for Climatology (Lisbon, April 1989) (L.J. Ogallo) and lectures presented at the training seminar in Muñoz, Philippines (14-24 November 1988) by N. Gbeckor-Kove] |
| WCAP - 8 | REPORT OF THE FIRST SESSION OF THE CCl WORKING GROUP ON CLIMATE AND URBAN AREAS INCLUDING BUILDING AND OTHER ASPECTS AND SOME RELATED PAPERS by Professors E. Jauregui and Shen Jianzhu, Members of the Working Group |
| WCAP - 9 | REPORT OF THE EXPERT MEETING ON CLICOM CLIMATE APPLICATIONS (INCLUDING CARS), Geneva, 6-10 November 1989 |
| WCAP - 10 | URBAN DESIGN IN DIFFERENT CLIMATES by B. Givoni, University of California, U.S.A. |
| WCAP - 11 | FIFTH PLANNING MEETING ON WORLD CLIMATE PROGRAMME - WATER, Laxenburg, Austria, 30 April - 4 May 1990 |
| WCAP - 12 | IMPACT POSSIBLE DES CHANGEMENTS CLIMATIQUES A VENIR SUR LES RESSOURCES EN EAU DES REGIONS ARIDES ET SEMI-ARIDES, par Jacques Sircoulon, ORSTOM, Paris, France, June 1990 |
| WCAP - 13 | INFORMATION ON METEOROLOGICAL EXTREMES FOR THE DESIGN AND OPERATION OF ENERGY SYSTEMS by G.A. McKay, Consulting climatologist, Canada, September 1990 |
| WCAP - 14 | EXTREMES AND DESIGN VALUES IN CLIMATOLOGY by Tibor Faragó, Hungarian Meteorological Service, Budapest, Hungary and Richard W. Katz, National Center for Atmospheric Research, Boulder, U.S.A. |